NUWC-NPT Technical Document 10,785
12 December 1994

Adaptive Block Sequential Detection of Abrupt Signal Changes

D. A. Abraham
Submarine Sonar Department

Naval Undersea Warfare Center Division
Newport, Rhode Island

PREFACE

This document was prepared as a Ph.D. Dissertation for a Doctor of Philosophy of Electrical Engineering at the University of Connecticut, Storrs, Connecticut. The major adviser for this work was Professor Charles H. Knapp. Associate advisers were Professor Peter K. Willett and Professor Yaakov Bar-Shalom.

The technical reviewer for this document was James Nuttall (Code 2121).

The author gratefully acknowledges James Nuttall for his careful review of this document. This work was partially sponsored by the Office of Naval Research.

Reviewed and Approved: 12 December 1994

R. J. Martin
Head: Submarine Sonar Department *(acting)*

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	12 December 1994	Final

4. TITLE AND SUBTITLE

Adaptive Block Sequential Detection of Abrupt Signal Changes

5. FUNDING NUMBERS

PR A60070

6. AUTHOR(S)

Douglas A. Abraham

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Undersea Warfare Center Detachment
39 Smith Street
New London, Connecticut 06320-5594

8. PERFORMING ORGANIZATION REPORT NUMBER

TD 10,785

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Ph.D dissertation for a Doctor of Philosophy of Electrical Engineering, University of Connecticut.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Estimation of the time of change of a parameter governing an observed sequence of independent multivariate random data samples is considered when the magnitude of the change is unknown and there exist unknown nuisance parameters that may be nonstationary. The particular application considered is the detection of the onset of a narrowband signal with unknown amplitude and phase at an array of sensors for radar or sonar processing in the presence of unknown, nonstationary, spatially distinct interferences hindering detection.

The solution proposed in this dissertation entails segmenting the multivariate data into nonoverlapping blocks, from which univariate statistics are formed that are invariant to the unknown interference parameters. The log likelihood ratio for a specific signal-to-interference ratio (SIR) and the locally optimal nonlinearities are applied to two block level statistics formed by generalized likelihood ratio (GLR) methods for data compression prior to submission to Page's test for rapid detection of the change. Once Page's test has determined that a change has occurred at the block level, the estimate of the change time is improved by post block processing in the form of a maximum likelihood estimator for the change time.

Application of the locally optimal nonlinearity to the block level statistic formed from a GLR for the unknown signal and interference parameters achieved the best asymptotic performance. Increasing the dimension of the

14. SUBJECT TERMS		15. NUMBER OF PAGES
Signal-to-Interference Ratio Segmenting Data	Random Data Samples	133
		16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

13. (Cont'd)

multivariate data applied to the detector caused a corresponding increase in the SIR required to achieve a specified asymptotic performance. Submitting the full array data to the detector provides the most interference suppression, resulting in the largest achieved SIR, but also the largest required SIR for a specified performance. Reducing the adaptive dimension of the array data by a constant matrix preprocessor, such as a beamspace preprocessor, provides a trade-off between the increased required SIR and the reduced interference suppression. Curves of the SIR required to achieve a specified performance as a function of the dimension of the data submitted to the detector provide a method for choosing the preprocessor dimension that maximizes the asymptotic performance for a specific interference.

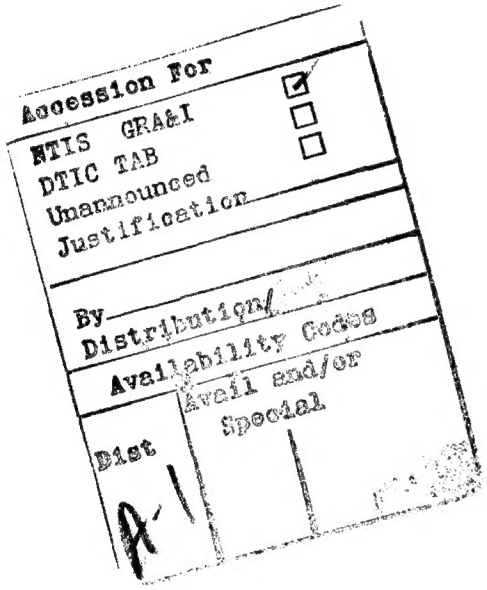

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

The general problem to be considered in this dissertation is the rapid determination of the time that a particular distributional characteristic or parameter of a sequence of independent multivariate random data samples changes. This topic has been studied for applications in typical array processing fields such as radar, sonar and the processing of biomedical signals, as well as other fields such as fault detection in linear systems, signal segmentation in speech processing, and change point problems in statistics. The problem is described as estimating the time, P, that the change occurs under the statistical model

$$X_i \sim \begin{cases} \mathrm{F}_H(\theta_{Hi}) & i < P \\ \mathrm{F}_K(\theta_{Ki}) & i \geq P \end{cases} . \tag{1}$$

That is, X_i, the data for time sample i, is independently drawn from the distribution $\mathrm{F}_H(\theta_{Hi})$ for $i < P$ and from the distribution $\mathrm{F}_K(\theta_{Ki})$ for $i \geq P$ where

the cumulative distribution functions, F_H and F_K, respectively depend on time dependent parameter vectors θ_{Hi} and θ_{Ki}.

Page [1] provides a sequentially based method for the quick detection of the change from one distributional model to another *when each distribution is exactly specified and stationary.* Page's test may be best described as consecutive applications of Wald's sequential probability ratio test (SPRT). The SPRT's are designed to decide between two stationary, known hypotheses; the signal absent (null) hypothesis and the signal present (alternative) hypothesis. These assumptions imply that the parameter vectors of equation (1) are known and constant, $\theta_{Hi} = \theta_H$ and $\theta_{Ki} = \theta_K$. Appropriately designed SPRT's will eventually terminate, choosing the null hypothesis with a certain design probability, when there is no signal present. When this happens, Page's test restarts the SPRT assuming that the signal is not present yet. When the signal does occur, the current SPRT or a following SPRT will terminate at the alternative hypothesis with a certain design probability, thus, detecting the signal. Performance parameters, differing from the Neyman-Pearson concepts of detection and false alarm probabilities (or Type I and II error probabilities), are the average number of samples between false alarms and the average number of samples before detection. Page's test and Wald's SPRT are designed for simple or point hypothesis testing, meaning that the statistical distribution of the data is known exactly under both the null

and alternative hypotheses. It is not, in general, clear how to proceed when either or both distributions are not completely characterized, yielding composite hypotheses.

One application of this problem involves the detection of the onset of a signal received at an array of sensors in a background of interferences and noise. The signal may represent a radar or sonar target return, seismic disturbance or the failure of a mechanical system. In frequency domain array processing applications, the data is usually modeled as a complex Gaussian random vector [2] which is completely characterized by its mean vector and covariance matrix. Before the signal arrives, the data is assumed to have a zero mean vector. After the signal arrives, under the deterministic signal model, the mean vector is characterized by an unknown complex scalar (representing the magnitude and phase of a sinusoid arriving at the array) multiplied by the array steering or replica vector. The array steering vector is assumed to be known and represents the relative amplitude and phase effects of a signal being propagated from a hypothesized position to the array sensors. In both cases, the covariance matrix of the data, representing the interfering signals and noise, is unknown and may be slowly varying. The simplest graphical example of this *shift in mean* scenario is the univariate or single sensor case shown in figure 1 with real rather than complex data.

Research related to the array signal onset detection problem has congregated into three areas: failure detection in linear dynamical systems, the application of Page's test to simplified versions of the problem, and sliding block techniques

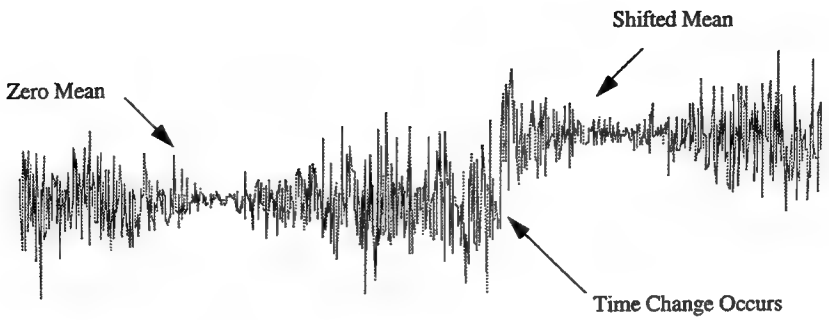

Figure 1: Example of a data sequence from a shift in mean type distributional change for univariate Gaussian data with non-stationary variance.

assuming short term stationarity with Neyman-Pearson based testing for the presence of a signal. Basseville and Nikiforov [3] have recently presented a compilation of many existing algorithms for change detection, however, interference covariance matrices are treated as identity matrices or assumed to be known. There exists a substantial amount of research in the failure detection area as discussed by Kerr [4] and Basseville and Benveniste [5], with several sequential type solutions. A linear time-varying state model, with an additional additive impulsive component representing the failure or change in state, is typically used in this area as described by Brumback [6] and Kerr [7]. This is a particularly appealing model for radar or sonar applications for it may accurately represent the non-stationarity of the interference and noise structure. However, since complete knowledge of the time-varying model parameters is assumed, it is not appropriate for the unknown, slowly changing interference and noise case. Fortunately, there exist similarities including unknown fault levels and the difficulties involved in choosing detection thresholds.

As previously mentioned, Page's test provides a method for detecting the change from one known distribution to another known distribution. Clearly the unknown complex signal amplitude and the unknown interference covariance structure, as well as its non-stationarity, will cause substantial problems in applying this method. Analysis to date in this area includes applying Page's test to univariate (single sensor) cases with known interference and unknown signal characteristics. For this univariate case, Dyson [8], Broder [9], and Stahl [10] have considered locally optimal test statistics to avoid the unknown signal strength problem and have assumed ideal normalization, which is equivalent to complete knowledge of the interference structure. Blostein [11] has considered the univariate, ideal normalization case with a non-stationary signal of known form and unknown amplitude. A Bayesian formulation of the problem involves applying a prior density to the unknown signal strength as described in Basseville and Nikiforov [3]. It seems that applying Page's test to the multivariate, unknown interference case has not been considered.

In resolving the interference non-stationarity problem, a short block of data that is assumed to have stationary interference is often analyzed. Here the multivariate case has been addressed as well as that of an unknown interference covariance structure. Generalized likelihood methods are used to form test statistics that are compared to thresholds for deciding signal presence in the Neyman-Pearson fashion (attempting to maximize the detection probability for a fixed false alarm rate). The block is then shifted along the data sequence, continually

testing for signal presence. Typically a trailing block that is assumed to contain no signal is used to aid in the estimation of the interference covariance structure. A statistical analysis of this concept was first performed by Reed, Mallet and Brennan [12]. A detector having a constant false alarm rate (CFAR) property was first introduced by Kelly [13], followed by Robey [14] and Chen and Reed [15]. The detector based on the work of Reed, Mallet and Brennan [12] does not have the CFAR property. Therefore, detection thresholds will depend on the unknown interference covariance structure. The latter developments [13]-[15] provide methods for choosing thresholds which makes implementing the detector feasible. The CFAR property is the result of forming test statistics that have probability density functions (PDF's) that do not depend on the interference covariance structure when no signal is present. Unfortunately, the PDF's of these test statistics, and thus detector thresholds, are often very difficult to derive and evaluate.

In summarizing relevant existing research it is found that sequential techniques utilizing Page's test have been developed for known, stationary interference covariance structures in the univariate case allowing an unknown signal amplitude and phase. Those methods that deal with the multivariate, unknown interference covariance are based around non-sequential, Neyman-Pearson testing using statistics formed from short blocks of data over which the interference is assumed to be stationary.

This research proffers a technique that involves segmenting the incoming data sequence into non-overlapping blocks of equal size small enough that the data within each block may be assumed to be stationary. From each block, a univariate statistic is formed in such a fashion that its probability density function does not depend on the interference covariance matrix when no signal is present. This property is known as ancillarity and insures that any detector based on this statistic will have a constant false alarm rate. The ancillarity of the block level statistic to the interference covariance structure clearly solves the unknown interference covariance problem at the expense of time resolution. The ancillarity also deals with the non-stationarity of the interference from block to block. Indirect dependence on the interference structure is allowed, when a signal is present, in the form of a scalar quantity, the signal-to-interference power ratio (SIR). There exists no *optimal* method for forming the block level statistic, so two approaches are considered that utilize generalized likelihood ratio techniques.

From a microscopic point of view, the problem has now become a simple versus composite hypothesis test, in that the distribution of the block level statistic is known explicitly when no signal is present and only depends on the scalar SIR otherwise. This situation now resembles what Dyson [8] considered and is treated by using a locally optimal detector non-linearity. An alternative approach is to form a test based on a design signal-to-interference ratio in the hopes that the test will work well for a wider range of SIR. These two methods provide detector functions or non-linearities that are applied to the block level statistic prior to

submission to Page's test. Page's test will, with some delay, determine when the signal has occurred at the block level. Post block processing, in the form of a maximum likelihood estimate, provides higher resolution estimation of the onset time. A block diagram describing the proposed algorithm can be found in figure 2.

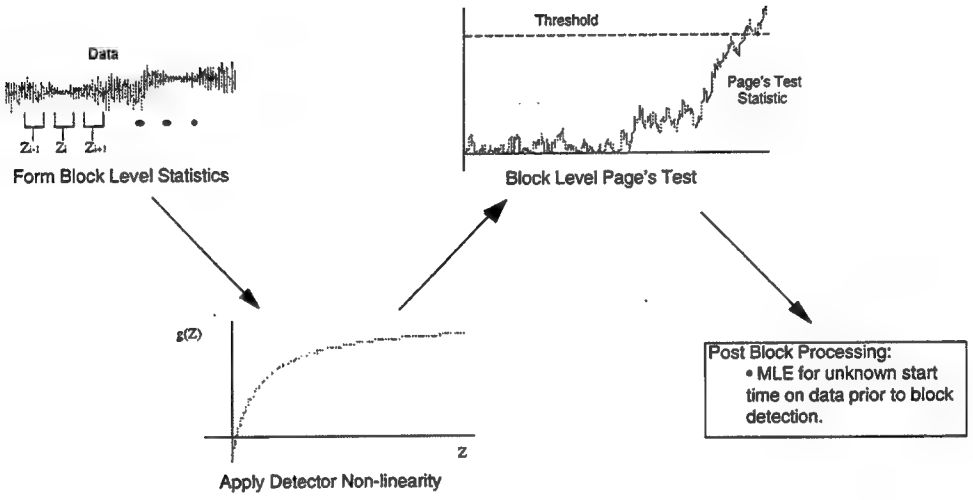

Figure 2: Block diagram of proposed algorithm.

The basic problem addressed in this dissertation is the application of Page's test to the multivariate unknown, slowly varying interference covariance matrix and unknown complex signal amplitude problem. Although this is a specific problem, the solution is applicable to a particular subset of the general problem as stated in equation (1): those problems that have common unknown (nuisance) parameters before and after the change takes place, subject to the existence of block statistics that are ancillary for the nuisance parameters under the null hypothesis and depend on at most an unknown scalar *strength* parameter under the alternative hypothesis.

This dissertation is organized as follows: Background information on Page's test, performance measures, and the choice of the detector non-linearity are discussed in chapter 2. Chapter 3 contains two approaches for forming the block level statistic, the derivation of each statistic, and a statistical description in the form of conditional probability density functions. The locally optimal detector non-linearity derivation and optimal methods of choosing a bias for Page's test follow in chapter 4. The performance of the detectors is explored for varying SIR, number of sensors, and block size in chapter 5. The post block processing algorithm is derived and evaluated in chapter 6. A simulation, coupling the block detection utilizing Page's test to the post block processing, for a stationary and a non-stationary unknown interference scenario is found in chapter 7, followed by concluding remarks and recommendations in chapter 8. Several theorems are used throughout this dissertation with their statements and proofs found in appendix A. A method for approximating the non-central Fisher's f probability density function is described in appendix B. Appendix C contains a summary of the probability density functions and selected properties of the random variables used in this dissertation.

Chapter 2

Page's Test Background

The detector structure proposed in this dissertation utilizes Page's test [1] to determine the time that a distributional parameter of a sequence of random data samples changes from following a null hypothesis (H) to following an alternative hypothesis (K). Page's test uses, as a detector statistic, the difference between the current value of a cumulative sum statistic and the minimum value over all past sums. The cumulative sum statistic is the sum of a detector non-linearity, g, applied to the data, X_i, at time sample i,

$$S_i = \sum_{k=1}^{i} g\left(X_k\right). \tag{2}$$

Page's test compares the detector statistic,

$$T_i = S_i - \min_{j \leq i} S_j, \tag{3}$$

to a threshold, h,

$$T_i \underset{H}{\overset{K}{\gtrless}} h, \tag{4}$$

to test for the alternative hypothesis. The detector statistic may alternatively be written in a recursive form,

$$T_i = \max\{0, T_{i-1} + g(X_i)\}. \tag{5}$$

The expected value of the detector non-linearity is required to be negatively biased under the null hypothesis and positively biased under the alternative,

$$\mathrm{E}[g|\theta_H] < 0 < \mathrm{E}[g|\theta_K], \tag{6}$$

where $\mathrm{E}[g|\theta_H]$ and $\mathrm{E}[g|\theta_K]$ respectively represent the expected value of the output of the detector non-linearity under the null and alternative hypotheses.

2.1 Performance Measures

The performance of Page's test, when applied to signal detection, is defined by the average number of samples between false detections when no signal is present, T, and the worst case average number of samples required before detection when a signal is present, D. The worst case condition is taken over all possible starting times for the signal and occurs whenever the statistic of equation (3) or (5) returns a value of zero. Broder [9] has shown that for large values of T, there is an exponential relationship between the average time between false alarms

and the average delay before detection. From this relationship an asymptotic performance measure is formed,

$$\eta = \lim_{h \to \infty} \frac{\log T}{D}, \tag{7}$$

where the natural logarithm is used and h is the threshold used in Page's test, which is monotonically increasing in T. As shown by Broder [9] and Willett [16], approximations to T and D may be found by using standard techniques of sequential analysis, resulting in

$$T = \frac{1 + ht(\theta_H) - e^{ht(\theta_H)}}{t(\theta_H)\,\mathrm{E}[g|\theta_H]} \tag{8}$$

$$\approx \frac{-e^{ht(\theta_H)}}{t(\theta_H)\,\mathrm{E}[g|\theta_H]} \tag{9}$$

and

$$D \approx \frac{h}{\mathrm{E}[g|\theta_K]}, \tag{10}$$

where $t(\theta_H)$ is the non-zero, unity root of the moment generating function of the detector non-linearity under the null hypothesis,

$$\mathrm{E}\left[e^{t(\theta_H)g}|\theta_H\right] = 1. \tag{11}$$

The approximations in (9) and (10) are valid for large values of the detector threshold, h, and are also, as seen in Broder [9], respectively lower and upper bounds. Applying these approximations to the asymptotic performance measure results in the approximation

$$\tilde{\eta} \approx \lim_{h \to \infty} \frac{\mathrm{E}[g|\theta_K]\left[ht(\theta_H) - \log\left(-t(\theta_H)\,\mathrm{E}[g|\theta_H]\right)\right]}{h}$$

$$= t(\theta_H)\,\mathrm{E}[g|\theta_K]$$

$$= \eta. \tag{12}$$

As mentioned by Broder [9], the approximation, η, is also a lower bound on the asymptotic performance measure of equation (7). In this dissertation, the approximation to Broder's asymptotic performance measure, η, will be termed the *asymptotic efficacy*.

From equation (9) the threshold required to implement Page's test, for a specified average time between false alarms, may be approximated by

$$\begin{aligned} h &\approx \frac{\log\left[-t(\theta_H)\,\mathrm{E}[g|\theta_H]\,T\right]}{t(\theta_H)} \\ &= \frac{\log T + \log\left[-t(\theta_H)\,\mathrm{E}[g|\theta_H]\right]}{t(\theta_H)}. \end{aligned} \tag{13}$$

2.2 Detector Non-Linearities

The detector non-linearity may be optimally chosen so that for a lower bounded average time between false alarms, $T \geq T_0$, the worst case average delay to detection is minimized. Lorden [17] proved this result in an asymptotic sense as $T_0 \to \infty$ for a log-likelihood ratio detector non-linearity,

$$l(x) = \log\left[\frac{f_X(x|\theta_K)}{f_X(x|\theta_H)}\right], \tag{14}$$

where $f_X(x|\theta_H)$ and $f_X(x|\theta_K)$ are respectively the probability density functions of the data under the null and alternative hypotheses. Moustakides [18] extended the result to the non-asymptotic case.

The locally optimal non-linearity for a scalar signal strength parameter, θ, was considered by Dyson [8],

$$g(x) = \frac{\left[\frac{\partial}{\partial\theta} f_X(x|\theta)\right]_{\theta=\theta_H}}{f_X(x|\theta_H)}, \tag{15}$$

and shown to be optimal for θ near θ_H. Dyson actually used a first order Taylor series expansion of the log-likelihood ratio about θ_H, including a bias term on the order of $(\theta - \theta_H)^2$, which results in a detector non-linearity that is still a function of the signal strength. Under the locally optimal or weak signal assumption, the bias term may be considered insignificant, and, if the signal strength is assumed to be constant, the result is a scale of equation (15). In order to apply (15) to Page's test, as seen in equation (6), a bias must be applied so that the mean of the detector non-linearity is negative when no signal is present. As seen in section 4.3, the bias may be chosen so as to minimize the signal strength required to achieve a desired asymptotic efficacy or to equivalently maximize the efficacy for a given signal strength.

Chapter 3

Block Level Statistics

The proposed solution to the data stationarity and unknown interference parameter problems involves segmenting the time series data into blocks of equal length. This allows each block of data to be compressed into a single univariate statistic having the predominant property of ancillarity with respect to the unknown interference parameters. A statistic formed from the random variable X with distribution $f_X(x|\theta)$ where θ is a parameter,

$$T = g(X), \tag{16}$$

is ancillary for θ if its distribution, $f_T(t)$, does not depend on θ [19]. The ancillarity avoids the stationarity problem in that, when no signal is present, a change in interference structure from one block to the next does not affect the distribution of the block level statistic.

In addition to the ancillarity of the block level test statistic to the interference parameters, it is desired that the statistic provide adequate detection performance when applied to Page's test for a wide range of signal powers. Ideally, the formation of the univariate test statistic from the block of multivariate data would be performed by optimizing some detection criteria related to Page's test, for instance, the asymptotic efficacy. Unfortunately, there is no known method for optimally dealing with the unknown interference parameters. The strategy adopted herein is to break the problem into two operations: first, compress the block level data into a univariate statistic that is ancillary for the unknown interference parameters, allowing dependence only through a signal related parameter, the signal-to-interference ratio (SIR); next, apply a non-linearity to the univariate statistic that, in some sense, optimizes the detection performance of Page's test.

In choosing a method of compressing the block data into a univariate statistic, the primary requirement is ancillarity with respect to the unknown interference parameters. Statistics that are invariant under a linear transformation of the array data under the null hypothesis provide ancillarity for the interference parameters as well as the potential for a *maximally* invariant statistic. Scharf [20] indicates that only considering decision rules that are invariant to such transformations is appropriate. The function that forms the univariate statistic from the block data is invariant under a linear transformation if the function has the same value for all linear transformations of original data [21] and [22]. The existence

of a maximally invariant statistic is not guaranteed and is, in general, difficult to test for. As an alternative to data compression under the premise of invariance, consider hypothesis testing in the presence of unknown parameters; the suboptimal, although usually very good, generalized likelihood ratio test (GLRT) is often used. This method of compressing block data into a univariate statistic may yield a statistic that is ancillary for the unknown interference parameters and is potentially good for detection in the Neyman-Pearson sense[1] .

When estimators of the unknown interference parameters exist in the same form when signal is present and absent, a block level statistic may be formed by substituting the estimates into a likelihood ratio test (LRT) or a GLRT for the unknown signal parameters where the interference parameters are assumed to be known. This method has been applied to adaptive array signal detection by Robey [14] and Chen and Reed [15], with the result termed an adaptive matched filter (AMF). In the following sections, the GLRT and the method involving the substitution of interference parameter estimates will be investigated for a deterministic signal with unknown complex amplitude. In particular, the function defining the compression of the block data will be derived and the block level statistic will be described statistically. The following sections apply these two methods of data compression to the unknown complex signal amplitude, unknown interference covariance matrix problem.

[1]The GLRT maximizes an upper bound on the best case detection probability while also bounding the worst case false alarm probability.

3.1 GLR Statistic

3.1.1 Derivation

The deterministic signal model assumes that the frequency domain array data is statistically represented by an N-variate complex Normal distribution,

$$\mathbf{x}_i \sim \mathcal{CN}_N\left(\theta\mathbf{d}, \mathbf{\Sigma}\right), \tag{17}$$

for $i = 1, \ldots, M$, where M is the block size, θ is the complex signal strength, $\mathbf{d}$ is a known mean vector, and $\mathbf{\Sigma}$ is the interference covariance matrix. The notation $\mathbf{x} \sim \mathcal{CN}_N\left(\mu, \mathbf{\Sigma}\right)$ indicates that the N-by-1 random vector $\mathbf{x}$ has a complex Normal distribution with mean μ and covariance matrix $\mathbf{\Sigma}$. Throughout this dissertation, bold lower case letters indicate vectors, bold uppercase letters indicate matrices, and the superscripts ^{T} and ^{H} respectively represent the transpose and complex conjugate and transpose operations.

The generalized likelihood ratio (GLR) statistic for the deterministic signal model has the following form,

$$\Lambda_{GLR} = \frac{\max\limits_{\theta\in\Theta, \mathbf{\Sigma}\in\Omega} f\left(\mathbf{X}|\theta, \mathbf{\Sigma}\right)}{\max\limits_{\mathbf{\Sigma}\in\Omega} f\left(\mathbf{X}|\mathbf{\Sigma}\right)}, \tag{18}$$

where

$$\Theta = \mathcal{C}^1 \tag{19}$$

and

$$\Omega = \left\{\mathbf{\Sigma} \in \mathcal{C}^{N\times N} : \mathbf{\Sigma} > \mathbf{0}, \mathbf{\Sigma} = \mathbf{\Sigma}^H\right\} \tag{20}$$

where $\mathcal{C}$ is the complex plane and $\mathbf{\Sigma} > \mathbf{0}$ indicates that the matrix $\mathbf{\Sigma}$ is positive definite. The likelihood function of the data in the block is

$$\begin{aligned} f(\mathbf{X}|\theta, \mathbf{\Sigma}) &= \prod_{i=1}^{M} \frac{1}{\pi^N |\mathbf{\Sigma}|} \mathrm{e}^{-(\mathbf{x}_i - \theta\mathbf{d})^H \mathbf{\Sigma}^{-1} (\mathbf{x}_i - \theta\mathbf{d})} \\ &= \left[\frac{1}{\pi^N |\mathbf{\Sigma}|} \mathrm{etr}\left(-\mathbf{\Sigma}^{-1} \frac{1}{M} \sum_{i=1}^{M} (\mathbf{x}_i - \theta\mathbf{d})(\mathbf{x}_i - \theta\mathbf{d})^H \right) \right]^M, \end{aligned} \quad (21)$$

as found in Goodman [23]. The function etr represents an exponential trace operator,

$$\mathrm{etr}(\mathbf{A}) = \mathrm{e}^{\mathrm{trace}(\mathbf{A})}, \quad (22)$$

and the N-by-M matrix $\mathbf{X}$ is the combined data of the block,

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \cdots \mathbf{x}_M]. \quad (23)$$

The GLR statistic for this problem may be found by a development similar to that of Kelly [13]. Kelly considered testing one data vector for the presence of a signal when independent secondary data with the same interference covariance structure was available to assist in estimation. The problem at hand results in a similar sufficient statistic probability density function. Proceeding, set

$$\mathbf{B}(\theta) = \frac{1}{M} \sum_{i=1}^{M} (\mathbf{x}_i - \theta\mathbf{d})(\mathbf{x}_i - \theta\mathbf{d})^H, \quad (24)$$

and substitute the deterministic signal model probability density function, (21), into the numerator of (18),

$$\max_{\theta \in \Theta, \mathbf{\Sigma} \in \Omega} f(\mathbf{X}|\theta, \mathbf{\Sigma}) = \max_{\theta \in \Theta, \mathbf{\Sigma} \in \Omega} \left[\frac{1}{\pi^N |\mathbf{\Sigma}|} \mathrm{etr}\left(-\mathbf{\Sigma}^{-1} \mathbf{B}(\theta) \right) \right]^M. \quad (25)$$

The maximization with respect to $\mathbf{\Sigma}$ is identical to that for determining the maximum likelihood estimate of an unknown covariance matrix for a Gaussian random vector with a zero mean. The specific result may be shown by equating the derivative, with respect to $\mathbf{\Sigma}$, of the quantity inside the brackets in equation (25) to zero,

$$\begin{aligned} \frac{\partial}{\partial \mathbf{\Sigma}} \frac{\operatorname{etr}\left(-\mathbf{\Sigma}^{-1}\mathbf{B}\left(\theta\right)\right)}{\pi^{N}\left|\mathbf{\Sigma}\right|} &= \frac{\operatorname{etr}\left[-\mathbf{\Sigma}^{-1}\mathbf{B}\left(\theta\right)\right]}{\pi^{N}\left|\mathbf{\Sigma}\right|}\left[\mathbf{\Sigma}^{-1}\mathbf{B}\left(\theta\right)\mathbf{\Sigma}^{-1}-\mathbf{\Sigma}\right] \\ &= \frac{\operatorname{etr}\left[-\mathbf{\Sigma}^{-1}\mathbf{B}\left(\theta\right)\right]}{\pi^{N}\left|\mathbf{\Sigma}\right|}\mathbf{\Sigma}^{-1}\left[\mathbf{B}\left(\theta\right)-\mathbf{\Sigma}\right]\mathbf{\Sigma}^{-1}, \end{aligned} \tag{26}$$

where the following matrix derivatives, as found in Scharf [20], are utilized

$$\frac{\partial}{\partial \mathbf{\Sigma}}\left|\mathbf{\Sigma}\right| = \left|\mathbf{\Sigma}\right|\left(\mathbf{\Sigma}^{-1}\right)^{H} \tag{27}$$

and

$$\frac{\partial}{\partial \mathbf{\Sigma}}\operatorname{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{A}\right) = \left(-\mathbf{\Sigma}^{-1}\mathbf{A}\mathbf{\Sigma}^{-1}\right)^{H}. \tag{28}$$

Clearly, the derivative of equation (26) equals zero when $\mathbf{\Sigma}$ is chosen to be

$$\hat{\mathbf{\Sigma}} = \mathbf{B}\left(\theta\right). \tag{29}$$

This in turn causes the numerator of the GLR statistic to become

$$\max_{\theta\in\Theta} f\left(\mathbf{X}|\theta,\hat{\mathbf{\Sigma}}\right) = \max_{\theta\in\Theta}\left[\frac{1}{\pi^{N}\left|\mathbf{B}\left(\theta\right)\right|}e^{-N}\right]^{M}. \tag{30}$$

The maximization over θ may be accomplished by minimizing the determinant of $\mathbf{B}\left(\theta\right)$. Proceeding, $\mathbf{B}\left(\theta\right)$ is written in terms of the sample mean and sample covariance matrix,

$$\begin{aligned} \mathbf{B}\left(\theta\right) &= \frac{1}{M}\sum_{i=1}^{M}\left[\left(\mathbf{x}_{i}-\bar{\mathbf{x}}\right)+\left(\bar{\mathbf{x}}-\theta\mathbf{d}\right)\right]\left[\left(\mathbf{x}_{i}-\bar{\mathbf{x}}\right)+\left(\bar{\mathbf{x}}-\theta\mathbf{d}\right)\right]^{H} \\ &= \mathbf{S}+\left(\bar{\mathbf{x}}-\theta\mathbf{d}\right)\left(\bar{\mathbf{x}}-\theta\mathbf{d}\right)^{H}, \end{aligned} \tag{31}$$

where

$$\bar{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_i \tag{32}$$

and

$$\mathbf{S} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^H. \tag{33}$$

As seen in Muirhead [24], the sample covariance matrix of equation (33), which is clearly hermitian, is positive definite with probability one if and only if $M > N$. Thus, a square root factorization may be applied where the square root matrix is also positive definite and invertible,

$$\begin{aligned} \mathbf{S} &= \mathbf{\Gamma}\mathbf{\Gamma}^H \\ \mathbf{\Gamma} &> 0. \end{aligned} \tag{34}$$

The determinant of $\mathbf{B}(\theta)$ may now be factored into

$$\begin{aligned} |\mathbf{B}(\theta)| &= \left|\mathbf{\Gamma}\left[\mathbf{I}_N + \mathbf{\Gamma}^{-1}(\bar{\mathbf{x}} - \theta\mathbf{d})(\bar{\mathbf{x}} - \theta\mathbf{d})^H(\mathbf{\Gamma}^H)^{-1}\right]\mathbf{\Gamma}^H\right| \\ &= \left|\mathbf{\Gamma}\mathbf{\Gamma}^H\right|\left[1 + (\bar{\mathbf{x}} - \theta\mathbf{d})^H(\mathbf{\Gamma}^H)^{-1}\mathbf{\Gamma}^{-1}(\bar{\mathbf{x}} - \theta\mathbf{d})\right] \\ &= |\mathbf{S}|\left[1 + (\bar{\mathbf{x}} - \theta\mathbf{d})^H\mathbf{S}^{-1}(\bar{\mathbf{x}} - \theta\mathbf{d})\right], \end{aligned} \tag{35}$$

where $\mathbf{I}_N$ is the N dimensional identity matrix. By completing the complex square in θ,

$$\begin{aligned} |\mathbf{B}(\theta)| &= |\mathbf{S}|\left[1 + \bar{\mathbf{x}}^H\mathbf{S}^{-1}\bar{\mathbf{x}} - \theta^*\mathbf{d}^H\mathbf{S}^{-1}\bar{\mathbf{x}} - \theta\bar{\mathbf{x}}^H\mathbf{S}^{-1}\mathbf{d} + \theta\theta^*\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}\right] \\ &= |\mathbf{S}|\left[1 + \bar{\mathbf{x}}^H\mathbf{S}^{-1}\bar{\mathbf{x}} - \frac{\left|\mathbf{d}^H\mathbf{S}^{-1}\bar{\mathbf{x}}\right|^2}{\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}} + \mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}\left|\theta - \frac{\mathbf{d}^H\mathbf{S}^{-1}\bar{\mathbf{x}}}{\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}}\right|^2\right], \end{aligned} \tag{36}$$

where the superscript * is the conjugation operator. Since $\mathbf{S}$ is positive definite, $\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}$ is greater than zero. Thus, $|\mathbf{B}(\theta)|$ is minimized by choosing

$$\hat{\theta} = \frac{\mathbf{d}^H\mathbf{S}^{-1}\bar{\mathbf{x}}}{\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}}. \tag{37}$$

Under the maximization with respect to θ, the numerator of the GLR statistic becomes

$$f(\mathbf{X}|\hat{\theta},\hat{\mathbf{\Sigma}}) = \frac{(\pi e)^{-MN}|\mathbf{S}|^{-M}}{\left[1+\bar{\mathbf{x}}^H\mathbf{S}^{-1}\bar{\mathbf{x}} - \frac{|\mathbf{d}^H\mathbf{S}^{-1}\bar{\mathbf{x}}|^2}{\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}}\right]^M}. \tag{38}$$

The denominator of the GLR statistic is exactly equation (30) with $\theta = 0$,

$$\max_{\mathbf{\Sigma}\in\Omega} f(\mathbf{X}|\mathbf{\Sigma}) = \left[\frac{1}{\pi^N|\mathbf{B}(0)|}e^{-N}\right]^M. \tag{39}$$

Setting $\theta = 0$ in equation (35) and substituting $|\mathbf{B}(0)|$ into (39) yields

$$f(\mathbf{X}|\hat{\mathbf{\Sigma}}) = \frac{(\pi e)^{-MN}|\mathbf{S}|^{-M}}{[1+\bar{\mathbf{x}}^H\mathbf{S}^{-1}\bar{\mathbf{x}}]^M}. \tag{40}$$

By combining (38) and (40), the resulting GLR statistic is

$$\Lambda_{GLR} = \left[\frac{1+\bar{\mathbf{x}}^H\mathbf{S}^{-1}\bar{\mathbf{x}}}{1+\bar{\mathbf{x}}^H\mathbf{S}^{-1}\bar{\mathbf{x}} - \frac{|\mathbf{d}^H\mathbf{S}^{-1}\bar{\mathbf{x}}|^2}{\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}}}\right]^M. \tag{41}$$

3.1.2 Conditional Distribution

The M^{th} root of the GLR statistic may equivalently be used in evaluating the conditional distribution, as it is a one-to-one transformation from Λ_{GLR},

$$T = (\Lambda_{GLR})^{\frac{1}{M}}. \tag{42}$$

The first step in statistically describing T for analysis is to form *whitened* data. This is accomplished by a full rank linear transformation of the block level data

so that the covariance matrix of the resulting transformed variables is the N dimensional identity matrix. The linear transformation is described by

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i, \tag{43}$$

for $i = 1, \ldots, M$. Because the transformation is full rank, it may additionally be defined to rotate the mean vector to a scaled unit vector with weight only in the first element,

$$\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T. \tag{44}$$

This is accomplished by choosing the transformation matrix as

$$\mathbf{A} = \mathbf{P}\boldsymbol{\Gamma}^{-1}, \tag{45}$$

where $\boldsymbol{\Gamma}$ is a square root of the data covariance matrix, $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^H, \tag{46}$$

and $\mathbf{P}$ is an orthogonal matrix chosen so that

$$\mathbf{A}\theta\mathbf{d} = \theta\mathbf{P}\boldsymbol{\Gamma}^{-1}\mathbf{d} = \beta\mathbf{e}_1, \tag{47}$$

where β satisfies

$$|\beta|^2 = |\theta|^2 \mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d} = s, \tag{48}$$

where s is defined as the adaptive signal-to-interference ratio. This definition of the SIR is from the beam output of a minimum variance distortionless response

adaptive beamformer when the interference structure is known. The whitened block data vectors are now distributed as

$$\mathbf{y}_i \sim \mathcal{CN}_N\left(\beta\mathbf{e}_1, \mathbf{I}_N\right). \tag{49}$$

The sample mean of the whitened data,

$$\begin{aligned} \bar{\mathbf{y}} &= \frac{1}{M}\sum_{i=1}^{M}\mathbf{y}_i \\ &= \mathbf{A}\bar{\mathbf{x}}, \end{aligned} \tag{50}$$

is distributed as

$$\bar{\mathbf{y}} \sim \mathcal{CN}_N\left(\beta\mathbf{e}_1, \frac{1}{M}\mathbf{I}_N\right). \tag{51}$$

The sample covariance matrix of the whitened data has the form

$$\mathbf{C} = \frac{1}{M}\sum_{i=1}^{M}\left(\mathbf{y}_i - \bar{\mathbf{y}}\right)\left(\mathbf{y}_i - \bar{\mathbf{y}}\right)^H \tag{52}$$

$$= \mathbf{A}\left[\frac{1}{M}\sum_{i=1}^{M}\left(\mathbf{x}_i - \bar{\mathbf{x}}\right)\left(\mathbf{x}_i - \bar{\mathbf{x}}\right)^H\right]\mathbf{A}^H$$

$$= \mathbf{A}\mathbf{S}\mathbf{A}^H. \tag{53}$$

As seen in Muirhead [24] for real Normal random vectors, the sample mean vector and the sample covariance matrix of complex Normal random vectors are independent and the sample covariance matrix is the scale of a complex Wishart distributed matrix,

$$M\mathbf{C} \sim \mathcal{CW}_N\left(M-1, \mathbf{I}_N\right). \tag{54}$$

The notation

$$\mathbf{X} \sim \mathcal{CW}_N\left(L, \mathbf{\Sigma}\right) \tag{55}$$

indicates that the N-dimensional random matrix $\mathbf{X}$ has a complex Wishart distribution with L degrees of freedom and scale matrix $\mathbf{\Sigma}$.

Applying this whitening transformation to equation (41), the M^{th} root of the GLR statistic becomes

$$T = \frac{1 + \bar{\mathbf{y}}^H \mathbf{C}^{-1} \bar{\mathbf{y}}}{1 + \bar{\mathbf{y}}^H \mathbf{C}^{-1} \bar{\mathbf{y}} - \frac{\left| \mathbf{e}_1^H \mathbf{C}^{-1} \bar{\mathbf{y}} \right|^2}{\mathbf{e}_1^H \mathbf{C}^{-1} \mathbf{e}_1}}. \tag{56}$$

Choosing the transformation to rotate all the weight of the mean vector into the first element of the transformed data vectors separates the data into a *signal* subset and a *non-signal* subset. The signal subset is comprised of the first element of each transformed data vector, while the non-signal subset contains the remaining data,

$$\mathbf{y}_i = \begin{bmatrix} y_{ai} \\ \\ \mathbf{y}_{bi} \end{bmatrix}, \tag{57}$$

where the subscript $_a$ represents the signal data and the subscript $_b$ represents the non-signal data.

The transformed data may be grouped, column-wise, into the N-by-M matrix

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_M \end{bmatrix}. \tag{58}$$

Applying the signal–non-signal segmentation to this matrix yields

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} y_{a1} & y_{a2} & \cdots & y_{aM} \\ \\ \mathbf{y}_{b1} & \mathbf{y}_{b2} & \cdots & \mathbf{y}_{bM} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{y}_a^T \\ \\ \mathbf{Y}_b \end{bmatrix}, \end{aligned} \tag{59}$$

where signal data is grouped into an M-by-1 vector,

$$\mathbf{y}_a = \begin{bmatrix} y_{a1} & y_{a2} & \cdots & y_{aM} \end{bmatrix}^T, \tag{60}$$

and non-signal data is grouped into an $N-1$-by-M matrix,

$$\mathbf{Y}_b = \begin{bmatrix} \mathbf{y}_{b1} & \mathbf{y}_{b2} & \cdots & \mathbf{y}_{bM} \end{bmatrix}. \tag{61}$$

Applying this segmentation to the transformed data sample mean vector and sample covariance matrix yields respectively,

$$\bar{\mathbf{y}} = \begin{bmatrix} \bar{y}_a \\ \\ \bar{\mathbf{y}}_b \end{bmatrix} \tag{62}$$

and

$$\mathbf{C} = \begin{bmatrix} c_{aa} & \mathbf{c}_{ba}^H \\ \\ \mathbf{c}_{ba} & \mathbf{C}_{bb} \end{bmatrix}. \tag{63}$$

As seen in (56), the inverse of the sample covariance matrix is required, and may be segmented as follows:

$$\mathbf{C}^{-1} = \mathbf{D} = \begin{bmatrix} d_{aa} & \mathbf{d}_{ba}^H \\ \\ \mathbf{d}_{ba} & \mathbf{D}_{bb} \end{bmatrix}. \tag{64}$$

The relationships between the partitioned sample covariance matrix and its inverse are easily derived and may be found, for example, in Muirhead [24],

$$d_{aa}^{-1} = c_{aa} - \mathbf{c}_{ba}^H \mathbf{C}_{bb}^{-1} \mathbf{c}_{ba} \tag{65}$$

$$\mathbf{d}_{ba} = -d_{aa} \mathbf{C}_{bb}^{-1} \mathbf{c}_{ba} \tag{66}$$

$$\mathbf{C}_{bb}^{-1} = \mathbf{D}_{bb} - d_{aa}^{-1} \mathbf{d}_{ba} \mathbf{d}_{ba}^H. \tag{67}$$

Applying the signal–non-signal segmentation to the quadratic and bilinear forms of equation (56) results in,

$$\mathbf{e}_1^H \mathbf{C}^{-1} \mathbf{e}_1 = d_{aa} \tag{68}$$

$$\mathbf{e}_1^H \mathbf{C}^{-1} \bar{\mathbf{y}} = d_{aa} \bar{y}_a + \mathbf{d}_{ba}^H \bar{\mathbf{y}}_b \tag{69}$$

$$\bar{\mathbf{y}}^H \mathbf{C}^{-1} \bar{\mathbf{y}} = d_{aa} |\bar{y}_a|^2 + \bar{y}_a \bar{\mathbf{y}}_b^H \mathbf{d}_{ba} + \bar{y}_a^* \mathbf{d}_{ba}^H \bar{\mathbf{y}}_b + \bar{\mathbf{y}}_b^H \mathbf{D}_{bb} \bar{\mathbf{y}}_b. \tag{70}$$

Completing the complex square for $\bar{y}_a$ in (70) yields the form

$$\begin{aligned} \bar{\mathbf{y}}^H \mathbf{C}^{-1} \bar{\mathbf{y}} &= d_{aa} \left| \bar{y}_a + d_{aa}^{-1} \mathbf{d}_{ba}^H \bar{\mathbf{y}}_b \right|^2 + \bar{\mathbf{y}}_b^H \left(\mathbf{D}_{bb} - d_{aa}^{-1} \mathbf{d}_{ba} \mathbf{d}_{ba}^H \right) \bar{\mathbf{y}}_b \\ &= d_{aa}^{-1} \left| d_{aa} \bar{y}_a + \mathbf{d}_{ba}^H \bar{\mathbf{y}}_b \right|^2 + \bar{\mathbf{y}}_b^H \mathbf{C}_{bb}^{-1} \bar{\mathbf{y}}_b. \end{aligned} \tag{71}$$

Substituting these forms into (56) results in

$$T = \frac{1 + \bar{\mathbf{y}}_b^H \mathbf{C}_{bb}^{-1} \bar{\mathbf{y}}_b + d_{aa} \left| \bar{y}_a + d_{aa}^{-1} \mathbf{d}_{ba}^H \bar{\mathbf{y}}_b \right|^2}{1 + \bar{\mathbf{y}}_b^H \mathbf{C}_{bb}^{-1} \bar{\mathbf{y}}_b}. \tag{72}$$

Subtracting one from each side and scaling the resulting numerator and denominator by $2M$ yields the form

$$W = T - 1 = \frac{2M \left(1 + \bar{\mathbf{y}}_b^H \mathbf{C}_{bb}^{-1} \bar{\mathbf{y}}_b \right)^{-1} \left| \bar{y}_a + d_{aa}^{-1} \mathbf{d}_{ba}^H \bar{\mathbf{y}}_b \right|^2}{2M d_{aa}^{-1}}. \tag{73}$$

Conditioned on observed values of the non-signal data, the numerator and denominator of W are independent Chi-squared random variables. The numerator is non-central with two degrees of freedom and the denominator is central with $2(M-N)$ degrees of freedom. This will be shown by forming the numerator and denominator of W as quadratic forms with the vector $\mathbf{y}_a^*$. Under certain conditions, the quadratic form involving either real or complex Gaussian random vectors has a non-central Chi-squared distribution. The results for the real

case are commonly known and may be found in Searle [25]. The results for the complex case, found in Theorem A.1 of appendix A, are similar.

In order to express the numerator and denominator of W as quadratic forms, the sample covariance matrix, $\mathbf{C}$, must be described in terms of the signal and non-signal data. Expanding the sum of equation (52) results in the form

$$\begin{aligned} \mathbf{C} &= \frac{1}{M}\sum_{i=1}^{M}\mathbf{y}_i\mathbf{y}_i^H - \bar{\mathbf{y}}\bar{\mathbf{y}}^H \\ &= \frac{1}{M}\mathbf{Y}\mathbf{Y}^H - \bar{\mathbf{y}}\bar{\mathbf{y}}^H. \end{aligned} \tag{74}$$

Applying the signal–non-signal segmentation yields

$$\mathbf{C} = \frac{1}{M}\begin{bmatrix} \mathbf{y}_a^T\mathbf{y}_a^* & \mathbf{y}_a^T\mathbf{Y}_b^H \\ \mathbf{Y}_b\mathbf{y}_a^* & \mathbf{Y}_b\mathbf{Y}_b^H \end{bmatrix} - \begin{bmatrix} \bar{y}_a\bar{y}_a^* & \bar{y}_a\bar{\mathbf{y}}_b^H \\ \bar{y}_a^*\bar{\mathbf{y}}_b & \bar{\mathbf{y}}_b\bar{\mathbf{y}}_b^H \end{bmatrix}. \tag{75}$$

Noting that the segmented sample mean vector of the transformed data may be expressed as linear combinations of the total data,

$$\bar{y}_a = \tfrac{1}{M}\mathbf{1}^T\mathbf{y}_a = \tfrac{1}{M}\mathbf{y}_a^T\mathbf{1} \tag{76}$$

$$\bar{\mathbf{y}}_b = \tfrac{1}{M}\mathbf{Y}_b\mathbf{1}\ , \tag{77}$$

where $\mathbf{1}$ is a column vector of ones, the elements of the sample covariance matrix may be expressed as

$$\begin{aligned} c_{aa} &= \frac{1}{M}\mathbf{y}_a^T\mathbf{y}_a^* - \frac{1}{M^2}\mathbf{y}_a^T\mathbf{1}\mathbf{1}^T\mathbf{y}_a^* \\ &= \frac{1}{M}\mathbf{y}_a^T\left(\mathbf{I}_M - \frac{1}{M}\mathbf{1}\mathbf{1}^T\right)\mathbf{y}_a^* \\ &= \frac{1}{M}\mathbf{y}_a^T\mathbf{P}_m\mathbf{y}_a^*, \end{aligned} \tag{78}$$

$$\begin{aligned}
\mathbf{c}_{ba} &= \frac{1}{M}\mathbf{Y}_b\mathbf{y}_a^* - \frac{1}{M}\bar{\mathbf{y}}_b\mathbf{1}^T\mathbf{y}_a^* \\
&= \frac{1}{M}\left(\mathbf{Y}_b - \bar{\mathbf{y}}_b\mathbf{1}^T\right)\mathbf{y}_a^* \\
&= \frac{1}{M}\mathbf{Y}_b\left(\mathbf{I}_M - \frac{1}{M}\mathbf{1}\mathbf{1}^T\right)\mathbf{y}_a^* \\
&= \frac{1}{M}\mathbf{Y}_b\mathbf{P}_m\mathbf{y}_a^*, \qquad (79) \\
\mathbf{C}_{bb} &= \frac{1}{M}\mathbf{Y}_b\mathbf{Y}_b^H - \frac{1}{M^2}\mathbf{Y}_b\mathbf{1}\mathbf{1}^T\mathbf{Y}_b^H \\
&= \frac{1}{M}\mathbf{Y}_b\left(\mathbf{I}_M - \frac{1}{M}\mathbf{1}\mathbf{1}^T\right)\mathbf{Y}_b^H \\
&= \frac{1}{M}\mathbf{Y}_b\mathbf{P}_m\mathbf{Y}_b^H, \qquad (80)
\end{aligned}$$

where $\mathbf{P}_m$ is a dimension M projection matrix with form

$$\mathbf{P}_m = \mathbf{I}_M - \frac{1}{M}\mathbf{1}\mathbf{1}^T. \qquad (81)$$

Defining

$$\begin{aligned}
\kappa_b &= \left(1 + \bar{\mathbf{y}}_b^H\mathbf{C}_{bb}^{-1}\bar{\mathbf{y}}_b\right) \\
&= \frac{1}{M}\mathbf{1}^T\left[\mathbf{I}_M + \frac{1}{M}\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\mathbf{1}, \qquad (82)
\end{aligned}$$

and applying equations (65)-(67) and (76)-(80), the numerator of W becomes,

$$\begin{aligned}
W_N &= 2M\kappa_b^{-1}\left|\bar{y}_a + d_{aa}^{-1}\mathbf{d}_{ba}^H\bar{\mathbf{y}}_b\right|^2 \\
&= 2M\kappa_b^{-1}\left|\bar{y}_a - \mathbf{c}_{ba}^H\mathbf{C}_{bb}^{-1}\bar{\mathbf{y}}_b\right|^2 \\
&= 2M\kappa_b^{-1}\left|\frac{1}{M}\mathbf{y}_a^T\left[1 - \mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\bar{\mathbf{y}}_b\right]\right|^2 \\
&= 2\left|\frac{1}{\sqrt{M\kappa_b}}\mathbf{y}_a^T\left[\mathbf{I}_M - \frac{1}{M}\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\mathbf{1}\right|^2 \\
&= 2\left|\mathbf{y}_a^T\mathbf{p}_b\right|^2 \\
&= 2\mathbf{y}_a^T\left(\mathbf{p}_b\mathbf{p}_b^H\right)\mathbf{y}_a^*, \qquad (83)
\end{aligned}$$

where

$$\mathbf{p}_b = \frac{1}{\sqrt{M\kappa_b}}\left[\mathbf{I}_M - \frac{1}{M}\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\mathbf{1}. \tag{84}$$

If W_N is to be a Chi-squared random variable, as seen in Theorem A.1, the product of the matrix in the quadratic form (i.e., $\mathbf{p}_b\mathbf{p}_b^H$) and the covariance matrix of the Gaussian random vector must be idempotent. The vector $\mathbf{y}_a$ is formed from the first elements of the transformed data vectors, $\mathbf{y}_i$, which are distributed as described in (49). Thus, the vector $\mathbf{y}_a^*$ is distributed as

$$\mathbf{y}_a^* \sim \mathcal{CN}_M\left(\beta^*\mathbf{1}, \mathbf{I}_M\right). \tag{85}$$

Since the covariance matrix of this Gaussian random vector is the identity matrix, if the matrices in the quadratic forms of the numerator and denominator of W are idempotent, the quadratic forms are Chi-squared distributed. Being the outer product of a vector with its conjugate transpose, the numerator quadratic form matrix has rank equal to one and is idempotent if the magnitude of the vector is one. Proceeding, the inner product of the vector, $\mathbf{p}_b$, with its conjugate transpose is shown to be equal to one,

$$\begin{aligned}
\mathbf{p}_b^H\mathbf{p}_b &= \frac{1}{M\kappa_b}\mathbf{1}^T\left[\mathbf{I}_M - \frac{1}{M}\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\mathbf{P}_m\right]\left[\mathbf{I}_M - \frac{1}{M}\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\mathbf{1} \\
&= \frac{1}{M\kappa_b}\mathbf{1}^T\left[\mathbf{I}_M + \frac{1}{M^2}\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\mathbf{1} \\
&= \frac{1}{M\kappa_b}\mathbf{1}^T\left[\mathbf{I}_M + \frac{1}{M}\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\mathbf{1} \\
&= 1.
\end{aligned} \tag{86}$$

This simplification requires noting that the vector, $\mathbf{1}$, is in the null space of the projection matrix $\mathbf{P}_m$,

$$\mathbf{P}_m\mathbf{1} \;=\; \left(\mathbf{1}^T\mathbf{P}_m\right)^T \;=\; \mathbf{0}\,, \tag{87}$$

and that, since $\mathbf{P}_m$ is a projection matrix,

$$\mathbf{P}_M^2 = \mathbf{P}_m. \tag{88}$$

As seen in Theorem A.1 the non-centrality parameter of W_N is a function of the mean vector of $\mathbf{y}_a^*$ and the matrix in the quadratic form of equation (83),

$$\begin{aligned}
\delta_N &= 2\left|\beta\right|^2 \mathbf{1}^T\mathbf{p}_b\mathbf{p}_b^H\mathbf{1} \\
&= 2\left|\beta\mathbf{1}^T\mathbf{p}_b\right|^2 \\
&= \frac{2}{M\kappa_b}\left|\beta\mathbf{1}^T\left[\mathbf{I}_M - \frac{1}{M}\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\mathbf{1}\right|^2 \\
&= 2\kappa_b^{-1}M\left|\beta\right|^2 \\
&= 2M\rho s,
\end{aligned} \tag{89}$$

where $\rho = \kappa_b^{-1}$. Thus, conditioned on the non-signal data, W_N is distributed as

$$W_N \sim \mathcal{X}_2^2\left(\delta_N\right). \tag{90}$$

The denominator of W is now placed into a quadratic form with the vector $\mathbf{y}_a^*$,

$$\begin{aligned}
W_D &= 2Md_{aa}^{-1} \\
&= 2M\left(c_{aa} - \mathbf{c}_{ba}^H\mathbf{C}_{bb}^{-1}\mathbf{c}_{ba}\right)
\end{aligned}$$

$$
\begin{aligned}
&= 2\mathbf{y}_a^T \left[\mathbf{P}_m - \frac{1}{M}\mathbf{P}_m \mathbf{Y}_b^H \mathbf{C}_{bb}^{-1} \mathbf{Y}_b \mathbf{P}_m\right] \mathbf{y}_a^* \\
&= 2\mathbf{y}_a^T \left(\mathbf{P}_m \left[\mathbf{I}_M - \frac{1}{M}\mathbf{Y}_b^H \mathbf{C}_{bb}^{-1} \mathbf{Y}_b\right] \mathbf{P}_m\right) \mathbf{y}_a^* \\
&= 2\mathbf{y}_a^T \mathbf{P}_b \mathbf{y}_a^*, \qquad (91)
\end{aligned}
$$

where

$$
\mathbf{P}_b = \mathbf{P}_m \left[\mathbf{I}_M - \frac{1}{M}\mathbf{Y}_b^H \mathbf{C}_{bb}^{-1} \mathbf{Y}_b\right] \mathbf{P}_m. \qquad (92)
$$

The matrix in the denominator quadratic form is shown to be idempotent by equating the matrix to its square,

$$
\begin{aligned}
\mathbf{P}_b^2 &= \mathbf{P}_m \left[\mathbf{I}_M - \frac{1}{M}\mathbf{Y}_b^H \mathbf{C}_{bb}^{-1} \mathbf{Y}_b\right] \mathbf{P}_m \left[\mathbf{I}_M - \frac{1}{M}\mathbf{Y}_b^H \mathbf{C}_{bb}^{-1} \mathbf{Y}_b\right] \mathbf{P}_m \\
&= \mathbf{P}_m - \frac{2}{M}\mathbf{P}_m \mathbf{Y}_b^H \mathbf{C}_{bb}^{-1} \mathbf{Y}_b \mathbf{P}_m + \frac{1}{M}\mathbf{P}_m \mathbf{Y}_b^H \mathbf{C}_{bb}^{-1} \left(\frac{1}{M}\mathbf{Y}_b \mathbf{P}_m \mathbf{Y}_b^H\right) \mathbf{C}_{bb}^{-1} \mathbf{Y}_b \mathbf{P}_m \\
&= \mathbf{P}_m \left[\mathbf{I}_M - \frac{1}{M}\mathbf{Y}_b^H \mathbf{C}_{bb}^{-1} \mathbf{Y}_b\right] \mathbf{P}_m \\
&= \mathbf{P}_b. \qquad (93)
\end{aligned}
$$

The non-centrality parameter of W_D is

$$
\begin{aligned}
\delta_D &= 2|\beta|^2 \mathbf{1}^T \mathbf{P}_b \mathbf{1} \\
&= 2|\beta|^2 \mathbf{1}^T \mathbf{P}_m \left[\mathbf{I}_M - \frac{1}{M}\mathbf{Y}_b^H \mathbf{C}_{bb}^{-1} \mathbf{Y}_b\right] \mathbf{P}_m \mathbf{1} \\
&= 0, \qquad (94)
\end{aligned}
$$

thus resulting in a central Chi-squared distribution. The degrees of freedom parameter,

$$
r_D = 2\mathrm{tr}\left(\mathbf{P}_b \mathbf{I}_M\right)
$$

$$
\begin{aligned}
&= 2\mathrm{tr}\left(\mathbf{P}_m\right) - \frac{2}{M}\mathrm{tr}\left(\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\mathbf{P}_m\right) \\
&= 2\mathrm{tr}\left(\mathbf{I}_M - \frac{1}{M}\mathbf{1}\mathbf{1}^T\right) - \frac{2}{M}\mathrm{tr}\left(\mathbf{Y}_b\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\right) \\
&= 2\left(M-1\right) - 2\mathrm{tr}\left(\mathbf{I}_{N-1}\right) \\
&= 2\left(M-N\right). \qquad (95)
\end{aligned}
$$

Thus, conditioned on the non-signal data, W_D is distributed as

$$W_D \sim \mathcal{X}^2_{2(M-N)}\left(0\right). \qquad (96)$$

The ratio of a non-central Chi-squared random variable scaled by its degrees of freedom to a central Chi-squared random variable scaled by its degrees of freedom is distributed as a singly non-central Fisher's f distribution when the two random variables are independent. As seen in Searle [25] and Theorem A.2 of Appendix A, quadratic forms of, respectively, a real or complex Gaussian vector, are pairwise independent if the $\mathbf{\Sigma}$ product of the matrices is zero, where $bSig$ is the covariance matrix of the Gaussian vector (which is the identity $\mathbf{A\Sigma B} = \mathbf{0}$, where $\mathbf{A}$ and $\mathbf{B}$ are the matrices involved in the quadratic forms).

Consider the $\mathbf{\Sigma}$ product of the matrices in the quadratic forms of W,

$$
\begin{aligned}
\mathbf{P}_b\mathbf{I}_M\mathbf{p}_b\mathbf{p}_b^H &\propto \mathbf{P}_m\left[\mathbf{I}_M - \frac{1}{M}\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\mathbf{P}_m\left[\mathbf{I}_M - \frac{1}{M}\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\mathbf{1}\mathbf{p}_b^H \\
&= \mathbf{P}_m\left[\mathbf{I}_M - \frac{1}{M}\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\left[\mathbf{P}_m\mathbf{1} - \frac{1}{M}\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\mathbf{1}\right]\mathbf{p}_b^H \\
&\propto \left[\mathbf{P}_m - \frac{1}{M}\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\right]\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\mathbf{1}\mathbf{p}_b^H \\
&= \left[\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\mathbf{1} - \frac{1}{M}\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\mathbf{1}\right]\mathbf{p}_b^H \\
&= \left[\mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\mathbf{1} - \mathbf{P}_m\mathbf{Y}_b^H\mathbf{C}_{bb}^{-1}\mathbf{Y}_b\mathbf{1}\right]\mathbf{p}_b^H
\end{aligned}
$$

$$= \mathbf{0}. \tag{97}$$

Thus, conditioned on the non-signal data, W_N and W_D are independent and W is the scale of a non-central Fisher's f distribution. The distributional parameters of the numerator and denominator of W only depend on the non-signal data through the non-centrality parameter of the numerator, equation (89), in the form of the variable ρ. The probability density function of ρ is determined as follows:

$$\begin{aligned}
\rho &= \frac{1}{1+\bar{\mathbf{y}}_b^H \mathbf{C}_{bb}^{-1} \bar{\mathbf{y}}_b} \\
&= \frac{\frac{2\bar{\mathbf{y}}_b^H \bar{\mathbf{y}}_b}{\bar{\mathbf{y}}_b^H (M\mathbf{C}_{bb})^{-1} \bar{\mathbf{y}}_b}}{\frac{2\bar{\mathbf{y}}_b^H \bar{\mathbf{y}}_b}{\bar{\mathbf{y}}_b^H (M\mathbf{C}_{bb})^{-1} \bar{\mathbf{y}}_b} + 2\bar{\mathbf{y}}_b^H (M\mathbf{I}) \bar{\mathbf{y}}_b} \\
&= \frac{U}{U+V},
\end{aligned} \tag{98}$$

where

$$U = \frac{2\bar{\mathbf{y}}_b^H \bar{\mathbf{y}}_b}{\bar{\mathbf{y}}_b^H (M\mathbf{C}_{bb})^{-1} \bar{\mathbf{y}}_b} \tag{99}$$

and

$$V = 2\bar{\mathbf{y}}_b^H (M\mathbf{I}) \bar{\mathbf{y}}_b. \tag{100}$$

Utilizing Theorem A.5 of appendix A, U is seen to have a central Chi-squared distribution with $2(M-N+1)$ degrees of freedom,

$$U \sim \mathcal{X}^2_{2(M-N+1)}, \tag{101}$$

and is independent of $\bar{\mathbf{y}}_b$, because the sample mean vector, $\bar{\mathbf{y}}_b$, and the sample covariance matrix, $\mathbf{C}_{bb}$, are independent and $M\mathbf{C}_{bb} \sim \mathcal{CW}_{N-1}(M-1, \mathbf{I}_{N-1})$.

From Theorem A.1 of appendix A, V has a central Chi-squared distribution with $2(N-1)$ degrees of freedom,

$$V \sim \mathcal{X}^2_{2(N-1)}, \tag{102}$$

because $\bar{\mathbf{y}}_b \sim \mathcal{CN}_{N-1}\left(\mathbf{0}, \frac{1}{M}\mathbf{I}_{N-1}\right)$ and $(M\mathbf{I}_{N-1})\left(\frac{1}{M}\mathbf{I}_{N-1}\right)$ is idempotent with trace equal to $N-1$. The central Chi-squared distribution results from the non-centrality parameter, the quadratic form $\mathbf{0}^H(M\mathbf{I})\mathbf{0}$, being zero. Since U and $\bar{\mathbf{y}}_b$ are independent, U and V are independent. Thus, using Theorem A.6 of Appendix A, (98) is beta distributed,

$$\rho \sim \beta eta\left(M-N+1, N-1\right), \tag{103}$$

with probability density function,

$$f(\rho) = \frac{\Gamma(M)}{\Gamma(N-1)\Gamma(M-N+1)}\rho^{M-N}(1-\rho)^{N-2}, \tag{104}$$

for $0 \leq \rho \leq 1$. Scaling the numerator and denominator of W by the degrees of freedom of their Chi-squared distributions yields a Fisher's f distribution. Thus, conditioned on ρ,

$$\begin{aligned} Z_{GLR} &= \frac{\frac{W_N}{2}}{\frac{W_D}{2(M-N)}} \\ &= (M-N)W \sim f_{2,2(M-N)}(2M\rho s), \end{aligned} \tag{105}$$

where the notation $X \sim f_{a,b}(c)$ implies that the random variable X has a singly non-central Fisher's f distribution with degrees of freedom parameters a and b and with non-centrality parameter c.

3.2 AMF Statistic

As described by Robey [14], the adaptive match filter (AMF) statistic is created by forming a GLR statistic for the unknown complex signal power under the assumption that the interference covariance matrix is known. An estimate of the interference covariance matrix that is statistically independent of the GLR statistic is then substituted into this form. Robey considered the same primary-secondary data scenario as Kelly [13], as mentioned in section 3.1.1, where the secondary data is assumed to contain interference only and to be independent of the primary data. In the sequential block problem, it is seen that the GLR statistic derived in the above fashion is a function of the sample mean vector. Thus, the sample covariance matrix, which is independent of the sample mean vector in the Gaussian case, may be used as the interference covariance matrix estimate. It should be noted that, in the case of the deterministic signal model, the sample covariance matrix is not the maximum likelihood estimator of the interference covariance matrix.

3.2.1 Derivation

The GLR statistic for the unknown signal power as a function of the assumed interference covariance matrix has the form

$$\Lambda_{AMF}(\boldsymbol{\Sigma}) = \frac{\max_{\theta \in \Theta} f(\mathbf{X}|\theta, \boldsymbol{\Sigma})}{f(\mathbf{X}|\boldsymbol{\Sigma})}. \tag{106}$$

From equation (21), the likelihood function of the data in the block may be described as

$$f(\mathbf{X}|\theta,\boldsymbol{\Sigma}) = \left[\frac{1}{\pi^N|\boldsymbol{\Sigma}|}\text{etr}\left(-\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathbf{S}+\bar{\mathbf{x}}\bar{\mathbf{x}}^H-\theta\mathbf{d}\bar{\mathbf{x}}^H \\ -\theta^*\bar{\mathbf{x}}\mathbf{d}^H+\theta\theta^*\mathbf{d}\mathbf{d}^H\end{bmatrix}\right)\right]^M \tag{107}$$

$$= \left[\exp\left(-\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}\begin{bmatrix}\theta\theta^*-\theta\frac{\bar{\mathbf{x}}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}}{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}} \\ -\theta^*\frac{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}}{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}}+\left|\frac{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}}{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}}\right|^2\end{bmatrix}\right)\right.$$
$$\left.\times\frac{1}{\pi^N|\boldsymbol{\Sigma}|}\exp\left(\frac{\left|\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}\right|^2}{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}}-\bar{\mathbf{x}}^H\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}-\text{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{S}\right)\right)\right]^M$$

$$= \left[\exp\left(-\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}\left|\theta-\frac{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}}{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}}\right|^2\right)\right.$$
$$\left.\times\frac{1}{\pi^N|\boldsymbol{\Sigma}|}\exp\left(\frac{\left|\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}\right|^2}{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}}-\bar{\mathbf{x}}^H\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}-\text{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{S}\right)\right)\right]^M \tag{108}$$

Maximizing this with respect to θ occurs, as seen in (108), when

$$\hat{\theta}(\boldsymbol{\Sigma}) = \frac{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}}{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}}, \tag{109}$$

because the quadratic form, $\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}$, is always positive since $\boldsymbol{\Sigma}$ is a positive definite matrix. The likelihood function of the data in the block when no signal is present is exactly equation (107) when $\theta = 0$,

$$f(\mathbf{X}|\boldsymbol{\Sigma}) = \left[\frac{1}{\pi^N|\boldsymbol{\Sigma}|}\exp\left(-\text{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{S}\right)-\bar{\mathbf{x}}^H\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}\right)\right]^M. \tag{110}$$

Substituting (108) - (110) into the generalized likelihood ratio of (106) results in

$$\Lambda_{AMF}(\boldsymbol{\Sigma}) = \exp\left(\frac{\left|\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}\right|^2}{\mathbf{d}^H\boldsymbol{\Sigma}^{-1}\mathbf{d}}\right)^M. \tag{111}$$

The sample covariance matrix, $\mathbf{S}$, is now substituted into a one-to-one transformation of (111), resulting in the AMF statistic

$$Z_{AMF} = \log\left[\Lambda_{AMF}\left(\mathbf{S}\right)^{\frac{1}{M}}\right] = \frac{\left|\mathbf{d}^H\mathbf{S}^{-1}\bar{\mathbf{x}}\right|^2}{\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}}. \tag{112}$$

3.2.2 Conditional Distribution

In section 3.1.2, the distribution of the GLR statistic was determined conditioned on a set of non-signal data. The effects of the conditioning were represented as a Beta distributed random variable. The distribution of the AMF statistic may similarly be determined by conditioning on the sample covariance matrix. Using the whitening transformation described by equations (43) to (54), the AMF statistic may be written as

$$\begin{aligned} Z_{AMF} &= \frac{\left|\mathbf{e}_1{}^H\mathbf{C}^{-1}\bar{\mathbf{y}}\right|^2}{\mathbf{e}_1{}^H\mathbf{C}^{-1}\mathbf{e}_1} \\ &= \frac{\mathbf{e}_1{}^H\mathbf{C}^{-2}\mathbf{e}_1}{2M\mathbf{e}_1{}^H\mathbf{C}^{-1}\mathbf{e}_1} 2\bar{\mathbf{y}}^H\left(\frac{M\mathbf{C}^{-1}\mathbf{e}_1\mathbf{e}_1{}^H\mathbf{C}^{-1}}{\mathbf{e}_1{}^H\mathbf{C}^{-2}\mathbf{e}_1}\right)\bar{\mathbf{y}} \\ &= \frac{1}{\rho r} 2\bar{\mathbf{y}}^H\mathbf{P}\bar{\mathbf{y}}, \end{aligned} \tag{113}$$

where

$$\rho = \frac{\left(\mathbf{e}_1{}^H\mathbf{C}^{-1}\mathbf{e}_1\right)^2}{\mathbf{e}_1{}^H\mathbf{C}^{-2}\mathbf{e}_1}, \tag{114}$$

$$r = \frac{2M}{\mathbf{e}_1{}^H\mathbf{C}^{-1}\mathbf{e}_1}, \tag{115}$$

and

$$\mathbf{P} = M\frac{\mathbf{C}^{-1}\mathbf{e}_1\mathbf{e}_1{}^H\mathbf{C}^{-1}}{\mathbf{e}_1{}^H\mathbf{C}^{-2}\mathbf{e}_1}. \tag{116}$$

Writing r in the form

$$r = 2\frac{\mathbf{e}_1{}^H \mathbf{I}_N \mathbf{e}_1}{\mathbf{e}_1{}^H (M\mathbf{C})^{-1} \mathbf{e}_1} \tag{117}$$

and noting, as seen in (54), that $M\mathbf{C}$ has a complex Wishart distribution with $M-1$ degrees of freedom and the N dimensional identity matrix as a scale parameter and using Theorem A.5 of appendix A, r is seen to have a central Chi-squared distribution with $2(M-N)$ degrees of freedom,

$$r \sim \mathcal{X}^2_{2(M-N)}. \tag{118}$$

Based on the results of Reed, Mallet and Brennan [12], the parameter ρ (equation 29 of [12]) has a Beta distribution,

$$\rho \sim \beta eta\,(M-N+1, N-1)\,. \tag{119}$$

Both ρ and r are formed from the sample covariance matrix, leading one to believe that they may be correlated. Utilizing the data segmentation of equation (57) and the matrix partitioning of equations (63) and (64), the quadratic forms describing ρ and r may be expressed as

$$\mathbf{e}_1{}^H \mathbf{C}^{-1} \mathbf{e}_1 \;=\; d_{aa}, \tag{120}$$

and

$$\begin{aligned} \mathbf{e}_1{}^H \mathbf{C}^{-2} \mathbf{e}_1 &= d_{aa}^2 + \mathbf{d}_{ba}^H \mathbf{d}_{ba} \\ &= d_{aa}^2 \left(1 + \mathbf{c}_{ba}^H \mathbf{C}_{bb}^{-2} \mathbf{c}_{ba}\right). \end{aligned} \tag{121}$$

Substituting these forms into ρ and r results in

$$r = 2Md_{aa}^{-1}, \tag{122}$$

and

$$\rho = \left(1 + \mathbf{c}_{ba}^H \mathbf{C}_{bb}^{-2} \mathbf{c}_{ba}\right)^{-1}. \tag{123}$$

Using equation (65) and the first result of Theorem A.3 of appendix A, which states that $d_{aa}^{-1} = \mathbf{C}_{aa \cdot b}$ is independent of $\mathbf{c}_{ba}$ and $\mathbf{C}_{bb}$, it is seen that ρ and r are indeed independent.

Noting that $\bar{\mathbf{y}}$ has a covariance matrix equal to $\frac{1}{M}\mathbf{I}_N$, it is clear that its product with the matrix of the quadratic form in (113) is the outer product of a unit length vector with its conjugate transpose,

$$\begin{aligned} \frac{1}{M}\mathbf{I}_N\mathbf{P} &= \frac{\mathbf{C}^{-1}\mathbf{e}_1\mathbf{e}_1{}^H\mathbf{C}^{-1}}{\mathbf{e}_1{}^H\mathbf{C}^{-2}\mathbf{e}_1} \\ &= \left(\frac{\mathbf{C}^{-1}\mathbf{e}_1}{||\mathbf{C}^{-1}\mathbf{e}_1||_2}\right)\left(\frac{\mathbf{C}^{-1}\mathbf{e}_1}{||\mathbf{C}^{-1}\mathbf{e}_1||_2}\right)^H, \end{aligned} \tag{124}$$

and is thus idempotent. The notation

$$||\mathbf{x}||_2 = \sqrt{\mathbf{x}^H\mathbf{x}}, \tag{125}$$

is used to describe the two-norm of a vector. Applying Theorem A.1 of appendix A, the scaled quadratic form in equation (113) has a non-central Chi-squared distribution with two degrees of freedom,

$$2\bar{\mathbf{y}}^H\mathbf{P}\bar{\mathbf{y}} \sim \mathcal{X}_2^2(\delta), \tag{126}$$

and non-centrality parameter

$$\begin{aligned}\delta &= 2M\frac{\left(\mathbf{e}_1{}^H\mathbf{C}^{-1}\mathbf{e}_1\right)^2}{\mathbf{e}_1{}^H\mathbf{C}^{-2}\mathbf{e}_1}|\beta|^2 \\ &= 2M\rho s. \end{aligned} \tag{127}$$

The AMF statistic may now be described as the scale of a non-central Chi-squared random variable conditioned on independent Beta and central Chi-squared distributed random variables,

$$Z_{AMF} = \frac{1}{\rho r}V, \tag{128}$$

where

$$V \sim \mathcal{X}_2^2\left(2M\rho s\right), \tag{129}$$

and r and ρ, respectively, have distributions as described in equations (118) and (119). The conditioning on the central Chi-squared random variable, r, may be removed in the following manner:

$$\begin{aligned} f_Z\left(z|s,\rho\right) &= \mathrm{E}_r\left[f_Z\left(z|s,\rho,r\right)\right] \\ &= \mathrm{E}_r\left[\rho r f_V\left(\rho r z|s,\rho,r\right)\right] \\ &= \mathrm{E}_r\left[\rho r\sum_{k=0}^{\infty}\frac{e^{-\frac{\delta}{2}}\left(\frac{\delta}{2}\right)^k}{k!}\frac{\left(\rho r z\right)^k e^{-\rho r z/2}}{2^{k+1}\Gamma\left(k+1\right)}\right] \\ &= \rho\sum_{k=0}^{\infty}\frac{e^{-\frac{\delta}{2}}\left(\frac{\delta}{2}\right)^k}{k!}\frac{\left(\rho z\right)^k}{2^{k+1}\Gamma\left(k+1\right)}\mathrm{E}_r\left[r^{k+1}e^{-\rho r z/2}\right], \end{aligned} \tag{130}$$

where $f_V\left(v|s,\rho,r\right)$ is, as described in (129), a non-central Chi-squared distribution with two degrees of freedom and a non-centrality parameter $\delta = 2M\rho s$. The

resulting expectation over r may be simplified by utilizing the PDF of a Gamma random variable, requiring the general form

$$\begin{aligned}
\mathrm{E}_r\left[r^j e^{-tr}\right] &= \int_{r=0}^{\infty} r^j e^{-tr} f_r(r)\, dr \\
&= \int_{r=0}^{\infty} \frac{r^{M-N-1+j} e^{-\left(t+\frac{1}{2}\right)r}}{\Gamma(M-N)\, 2^{M-N}} dr \\
&= \left[\begin{array}{l} \frac{\Gamma(M-N+j)\left(t+\frac{1}{2}\right)^{-(M-N+j)}}{\Gamma(M-N)2^{M-N}} \\ \times \int_{r=0}^{\infty} \frac{r^{M-N+j-1} e^{-\left(t+\frac{1}{2}\right)r}}{\Gamma(M-N+j)\left(t+\frac{1}{2}\right)^{-(M-N+j)}} dr \end{array}\right] \\
&= \frac{\Gamma(M-N+j)\, 2^j}{\Gamma(M-N)\,(1+2t)^{M-N+j}}, \qquad (131)
\end{aligned}$$

where it is recognized that the integral in the third line of (131) is the integral of the probability density function of a standard Gamma random variable and, thus, equals one. Substituting this result into equation (130) yields,

$$\begin{aligned}
f_Z(z|s,\rho) &= \rho \sum_{k=0}^{\infty} \frac{e^{-\frac{\delta}{2}}\left(\frac{\delta}{2}\right)^k}{k!} \frac{\Gamma(M-N+k+1)}{\Gamma(M-N)\,\Gamma(k+1)} \frac{(\rho z)^k}{(1+\rho z)^{M-N+k+1}} \qquad (132) \\
&= \left(\frac{n_2}{n_1}\rho\right) \sum_{k=0}^{\infty} \frac{e^{-\frac{\delta}{2}}\left(\frac{\delta}{2}\right)^k}{k!} \frac{\Gamma\left(\frac{n_1+n_2}{2}+k\right)}{\Gamma\left(\frac{n_1}{2}+k\right)\Gamma\left(\frac{n_2}{2}\right)} \frac{n_2^{\frac{n_2}{2}} n_1^{\frac{n_1}{2}+k} \left(\frac{n_2}{n_1}\rho z\right)^{\frac{n_1}{2}+k-1}}{\left[n_2 + n_1\left(\frac{n_2}{n_1}\rho z\right)\right]^{\frac{n_1+n_2}{2}+k}},
\end{aligned}$$

where $n_1 = 2$ and $n_2 = 2(M-N)$. The last line of (132) may be recognized as the scale of a non-central Fisher's f distribution with degrees of freedom n_1 and n_2 and non-centrality parameter δ. Thus, conditioned on the random variable ρ, the AMF block level statistic may be described statistically by

$$Z_{AMF} = \frac{1}{(M-N)\,\rho} U, \qquad (133)$$

where

$$U \sim f_{2,2(M-N)}(2M\rho s), \qquad (134)$$

and, as previously described, ρ is Beta distributed.

Chapter 4

Detector Non-linearities

Once the data in the block has been compressed to a univariate statistic, it is possible to apply detector nonlinearities that may improve the sequential detection performance. As seen in sections 3.1.2 and 3.2.2, the conditional distributions of the GLR and AMF statistics are functions of a scalar quantity s, the signal-to-interference ratio. Assuming that the signal-to-interference ratio is known, the optimal detector non-linearity has the form of a log-likelihood ratio, as described in section 2.2. When the signal strength is not known, and weak signals are of interest, the locally optimal detector non-linearity is indicated as a reasonable solution, as described in section 2.2. In this chapter, the optimal detector is proposed for a specified value of signal-to-interference ratio, termed the design SIR, and the locally optimal detector is derived for the GLR and AMF statistics. In order to apply Page's test to the locally optimal detector non-linearities, a bias must be applied so that the mean of the detector non-linearity is negative when

no signal is present. It is seen that the bias may be chosen to equivalently minimize the signal-to-interference ratio required to achieve a prescribed asymptotic efficacy or to maximize the asymptotic efficacy for a prescribed SIR.

4.1 Optimal Detector Non-linearity for a Design SIR

Designing a detector that is optimal for a specified SIR may yield a test that has adequate performance over a wider range of SIR. The log-likelihood ratio form of the detector non-linearity requires the evaluation of the unconditioned probability density function (PDF) of the block level statistic under the null hypothesis (H), where $s = 0$, and under the alternative hypothesis (K) for $s = \tilde{s}$, where $\tilde{s}$ is the design signal-to-interference ratio,

$$l_{\tilde{s}}(z) = \log\left[\frac{f_Z(z|s=\tilde{s})}{f_Z(z|s=0)}\right], \tag{135}$$

where $f_Z(z|s)$ describes the unconditional distribution of the block level statistic, Z, for a signal-to-interference ratio equal to s.

The evaluation of the unconditioned PDF of the block level statistics requires evaluating a one dimensional integral of the non-central Fisher's f density function, which in turn is represented by an infinite summation. As described in [26] there exist several approximations to the cumulative distribution function of a non-central Fisher's f random variable. The most reasonable, in terms of computational requirements and accuracy, seems to be the Three-Moment Approximation proposed by Tiku [27]. Thus, the unconditional PDF's are evaluated using

Tiku's Three-Moment Approximation and numerical integration, as described in appendix B.

4.2 Locally Optimal Detector Non-linearity

The locally optimal detector non-linearity,

$$g(z) = \frac{\left[\frac{\partial}{\partial s} f_Z(z|s)\right]_{s=0}}{f_Z(z|s=0)}, \tag{136}$$

requires forming the partial derivative of the block level statistic PDF when a signal is present with respect to the signal-to-interference ratio followed by normalization by the block level statistic PDF when no signal is present.

4.2.1 GLR Statistic

The unconditional distribution of the GLR block level statistic may be described as

$$f_Z(z|s) = \mathrm{E}_\rho\left[f_Z(z|s,\rho)\right], \tag{137}$$

where, as described in (105), the PDF conditioned on knowledge of ρ, $f_Z(z|s,\rho)$, is a non-central Fisher's f distribution. Taking the partial derivative with respect to s inside the expectation over ρ and applying the chain rule for derivatives results in

$$\frac{\partial}{\partial s} f_Z(z|s) = \mathrm{E}_\rho\left[\frac{\partial}{\partial \delta} f_Z(z|s,\rho)\frac{\partial \delta}{\partial s}\right], \tag{138}$$

where $\delta = 2M\rho s$ is the non-centrality parameter of the Fisher's f density. The conditional PDF of the GLR statistic may be written as

$$f_Z(z|s,\rho) = \sum_{k=0}^{\infty} \frac{e^{-\frac{\delta}{2}}\left(\frac{\delta}{2}\right)^k}{k!} \frac{\Gamma(M-N+k+1)}{\Gamma(M-N)\Gamma(k+1)} \frac{(M-N)^{M-N} z^k}{(M-N+z)^{M-N+k+1}}, \quad (139)$$

from equation (105) and appendix C. Taking the partial derivative of the conditional PDF with respect to the non-centrality parameter, δ, requires the derivative forms

$$\frac{\partial}{\partial\delta} e^{-\frac{\delta}{2}}\delta^k = \begin{cases} e^{-\frac{\delta}{2}}\delta^{k-1}\left(k-\frac{\delta}{2}\right) & k \geq 1 \\ -\frac{e^{-\frac{\delta}{2}}}{2} & k = 0 \end{cases}, \quad (140)$$

which are evaluated at $\delta = 0$. Thus, $k = 0$ and $k = 1$ are the only non-zero terms in the infinite sum,

$$\begin{aligned} \left[\frac{\partial}{\partial\delta} f_Z(z|s,\rho)\right]_{s=0} &= \frac{(M-N)^{M-N}}{2\Gamma(M-N)} \begin{bmatrix} \frac{\Gamma(M-N+2)z}{(M-N+z)^{M-N+2}} \\ -\frac{\Gamma(M-N+1)}{(M-N+z)^{M-N+1}} \end{bmatrix} \\ &= \frac{(M-N)^{M-N+1}}{2(M-N+z)^{M-N+1}} \left[\frac{(M-N+1)z}{M-N+z} - 1\right] \\ &= \frac{(M-N)^{M-N+1}}{2(M-N+z)^{M-N+1}} \left(\frac{z-1}{1+\frac{z}{M-N}}\right). \end{aligned} \quad (141)$$

Combining this result with

$$\frac{\partial\delta}{\partial s} = 2M\rho \quad (142)$$

in equation (138) evaluated at $s = 0$ results in

$$\left[\frac{\partial}{\partial s} f_Z(z|s)\right]_{s=0} = \mathrm{E}_\rho\left[\frac{M\rho}{\left(1+\frac{z}{M-N}\right)^{M-N+1}} \left(\frac{z-1}{1+\frac{z}{M-N}}\right)\right]$$

$$= \frac{M\mathrm{E}_\rho[\rho]}{\left(1+\frac{z}{M-N}\right)^{M-N+1}}\left(\frac{z-1}{1+\frac{z}{M-N}}\right)$$
$$= \frac{M-N+1}{\left(1+\frac{z}{M-N}\right)^{M-N+1}}\left(\frac{z-1}{1+\frac{z}{M-N}}\right), \tag{143}$$

where the mean of ρ has the form (see appendix C)

$$\mathrm{E}_\rho[\rho] = \frac{M-N+1}{M}. \tag{144}$$

Setting $\delta = 0$ in (139) and taking the expectation over ρ, as described in (137), results in

$$f_Z(z|s=0) = \mathrm{E}_\rho\left[\frac{\Gamma(M-N+1)}{\Gamma(M-N)}\frac{(M-N)^{M-N}}{(M-N+z)^{M-N+1}}\right]$$
$$= \frac{1}{\left(1+\frac{z}{M-N}\right)^{M-N+1}}. \tag{145}$$

Here it is seen that the *unconditional* PDF of the GLR block level statistic, when no signal is present, is a central Fisher's f distribution. Now, forming the locally optimal non-linearity of equation (136) yields

$$g(z) = (M-N+1)\frac{z-1}{1+\frac{z}{M-N}}. \tag{146}$$

As seen in Broder [9], the asymptotic efficacy is invariant to scale changes of the detector non-linearity. Thus, the non-linearity of equation (146) may be scaled so that it approaches one as z approaches infinity. This results in the form

$$g_{GLR}(z) = \frac{z-1}{z+M-N}. \tag{147}$$

4.2.2 AMF Statistic

The unconditional probability density function for the AMF block level statistic may be described as in equation (137). The locally optimal detector nonlinearity may then be found by evaluating the partial derivative of the conditional distribution of the block level statistic as described in equation (138). The conditional distribution of the AMF statistic, as seen in equation (132), is a scaled non-central Fisher's f distribution. Taking the partial derivative with respect to the non-centrality parameter and setting it equal to zero yields results similar to that of the previous section where the $k = 0$ and $k = 1$ terms are the only non-zero terms in the summation,

$$\begin{aligned}\left[\frac{\partial}{\partial s} f_Z\left(z|s,\rho\right)\right]_{s=0} &= \frac{\rho}{2\Gamma\left(M-N\right)}\left[\frac{\Gamma\left(M-N+2\right)\rho z}{\left(1+\rho z\right)^{M-N+2}} - \frac{\Gamma\left(M-N+1\right)}{\left(1+\rho z\right)^{M-N+1}}\right] \\ &= \frac{\left(M-N\right)}{2}\left[\frac{\left(M-N+1\right)z\rho^2}{\left(1+\rho z\right)^{M-N+2}} - \frac{\rho}{\left(1+\rho z\right)^{M-N+1}}\right]. \quad (148)\end{aligned}$$

Combining this result with equation (142) in equation (138) evaluated at $s = 0$ results in

$$\begin{aligned}\left[\frac{\partial}{\partial s} f_Z\left(z|s\right)\right]_{s=0} &= M\left(M-N\right)\left[\begin{array}{c}\left(M-N+1\right)z\mathrm{E}_\rho\left[\frac{\rho^3}{\left(1+\rho z\right)^{M-N+2}}\right] \\ -\mathrm{E}_\rho\left[\frac{\rho^2}{\left(1+\rho z\right)^{M-N+1}}\right]\end{array}\right] \\ &= M\left(M-N\right)\left[\left(M-N+1\right)z h_{3,2}\left(z\right) - h_{2,1}\left(z\right)\right], \quad (149)\end{aligned}$$

where

$$\begin{aligned}h_{i,j}\left(z\right) &= \mathrm{E}_\rho\left[\frac{\rho^i}{\left(1+\rho z\right)^{M-N+j}}\right] \\ &= \int_{\rho=0}^{1} \frac{\Gamma\left(M\right)\rho^{M-N+i}\left(1-\rho\right)^{N-2}}{\Gamma\left(N-1\right)\Gamma\left(M-N+1\right)\left(1+\rho z\right)^{M-N+j}} d\rho. \quad (150)\end{aligned}$$

By setting s, and therefore δ, equal to zero in the conditional PDF of the AMF statistic found in equation (132), and taking the expectation with respect to ρ, the unconditional PDF when no signal is present is formed,

$$\begin{aligned} f_Z(z|s=0) &= (M-N)\,\mathrm{E}_\rho\left[\frac{\rho}{(1+\rho z)^{M-N+1}}\right] \\ &= (M-N)\,h_{1,1}(z). \end{aligned} \tag{151}$$

Placing equations (149) and (151) into equation (136), the AMF locally optimal non-linearity is formed,

$$g(z) = M(M-N+1)\,z\frac{h_{3,2}(z)}{h_{1,1}(z)} - M\frac{h_{2,1}(z)}{h_{1,1}(z)}. \tag{152}$$

As with the GLR locally optimal non-linearity, the AMF non-linearity is scaled so that its limit is one as z goes to infinity. It is first recognized that

$$\begin{aligned} \lim_{z\to\infty} z^{M-N+j}h_{i,j}(z) &= \lim_{z\to\infty}\mathrm{E}\left[\frac{\rho^i}{\left(\rho+\frac{1}{z}\right)^{M-N+j}}\right] \\ &= \mathrm{E}\left[\rho^{-(M-N-i+j)}\right]. \end{aligned} \tag{153}$$

Using the probability density function of the standard Beta distribution, the negative moments of ρ are seen to be

$$\mathrm{E}\left[\rho^{-K}\right] = \frac{\Gamma(M)\,\Gamma(M-N-K+1)}{\Gamma(M-K)\,\Gamma(M-N+1)} \tag{154}$$

when

$$K < M-N+1. \tag{155}$$

Applying equations (153) and (154) to the limit of equation (152) as $z \to \infty$ results in

$$\begin{aligned}\lim_{z\to\infty} g(z) &= \lim_{z\to\infty} M\left[(M-N+1)\frac{z^{M-N+2}h_{3,2}(z)}{z^{M-N+1}h_{1,1}(z)} - \frac{z^{M-N+1}h_{2,1}(z)}{z^{M-N+1}h_{1,1}(z)}\right] \\ &= M\left[(M-N+1)\frac{\mathrm{E}\left[\rho^{-(M-N-1)}\right]}{\mathrm{E}\left[\rho^{-(M-N)}\right]} - \frac{\mathrm{E}\left[\rho^{-(M-N-1)}\right]}{\mathrm{E}\left[\rho^{-(M-N)}\right]}\right] \\ &= \frac{M(M-N)}{N}, \end{aligned} \tag{156}$$

where

$$\begin{aligned}\frac{\mathrm{E}\left[\rho^{-(M-N-1)}\right]}{\mathrm{E}\left[\rho^{-(M-N)}\right]} &= \frac{\Gamma(M)\Gamma(2)\Gamma(N)\Gamma(M-N+1)}{\Gamma(N+1)\Gamma(M-N+1)\Gamma(M)\Gamma(1)} \\ &= \frac{1}{N}. \end{aligned} \tag{157}$$

Scaling equation (152) by the inverse of equation (156) results in the locally optimal non-linearity

$$g_{AMF}(z) = \frac{N}{M-N}\left[(M-N+1)\,z\frac{h_{3,2}(z)}{h_{1,1}(z)} - \frac{h_{2,1}(z)}{h_{1,1}(z)}\right]. \tag{158}$$

In general the locally optimal non-linearity for the AMF statistic is not strictly increasing. Thus, although it approaches one in the limit, the maximum may actually be greater than one. The $h_{i,j}(z)$ functions do not pose serious implementation problems as they may be easily evaluated using numerical integration and stored in tables.

4.3 Optimal Bias For Page's Test

As described in chapter 2 , when a detector non-linearity is applied to Page's test, the mean of the non-linearity must be negative when no signal is present and

positive otherwise. A log-likelihood ratio non-linearity satisfies this requirement, however, a locally optimal non-linearity may not. Stahl and Willett [10], following Dyson [8], have approached this problem by subtracting a bias proportional to the square of the signal strength. The proportionality constant is chosen by maximizing the fourth derivative of the asymptotic efficacy as the signal strength tends to zero.

An alternative solution, proposed herein, is to choose the subtractive bias that achieves a desired asymptotic efficacy with a minimum signal strength. In actual detector operation, a particular performance is typically desired and the strength of an observed signal is rarely known a priori. Thus, tuning the detector to yield the desired performance at a minimum signal strength is an appealing approach. It will be shown that this method is equivalent to choosing the bias that maximizes the efficacy given a fixed signal-to-interference ratio. Both the proposed method and that of Stahl and Willett result in a minimum detectable signal level, which is the consequence of a negative detector non-linearity mean when signals below the detectable level are present.

4.3.1 Derivation of the Optimal Bias

Let $g_\tau(x)$ be the detector non-linearity to be applied to Page's test,

$$g_\tau(x) = g_0(x) - \tau, \tag{159}$$

where τ represents the applied bias and $g_0(x)$ is the original non-linearity. It is assumed that the non-linearity is operating on a univariate test statistic modeled

by the random variable X with probability density function $f_X(x)$. Requiring the mean of $g_\tau(X)$ to be negative when no signal is present and positive otherwise yields

$$\mathrm{E}_0[g_\tau(X)] < 0 < \mathrm{E}_s[g_\tau(X)], \tag{160}$$

where the subscripts 0 and s on the expectations respectively represent the no signal and signal strength equal to s cases. Applying the bias form of the non-linearity results in

$$\mathrm{E}_0[g_0(X)] - \tau < 0 < \mathrm{E}_s[g_0(X)] - \tau, \tag{161}$$

or

$$\mathrm{E}_0[g_0(X)] < \tau < \mathrm{E}_s[g_0(X)]. \tag{162}$$

Equation (161) illuminates the minimum detectable signal phenomenon where, for a specific value of $\tau > 0$, there will be some s_{MDL} for which only $s > s_{MDL}$ will satisfy the right hand inequality of (161).

The detector non-linearity, $g_\tau(x)$, is now applied to the definition of the asymptotic efficacy (12),

$$\begin{aligned} \eta &= t_\tau \mathrm{E}_s[g_\tau(X)] \\ &= t_\tau(\mathrm{E}_s[g_0(X)] - \tau), \end{aligned} \tag{163}$$

which may be solved for the mean of the unbiased non-linearity when signal is present to yield

$$\mathrm{E}_s[g_0(X)] = \frac{\eta}{t_\tau} + \tau, \tag{164}$$

where t_τ is the non-zero, unity root of the moment generating function of the detector non-linearity with bias τ. Thus, t_τ satisfies

$$1 = \mathrm{E}_0\left[\mathrm{e}^{t_\tau g_\tau(X)}\right]. \tag{165}$$

Substituting in the bias form of the non-linearity results in

$$\begin{aligned} 1 &= \mathrm{E}_0\left[\mathrm{e}^{t_\tau(g_0(X)-\tau)}\right] \\ &= \mathrm{e}^{-t_\tau\tau}\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right], \end{aligned} \tag{166}$$

which may be solved for τ to yield

$$\tau = \frac{\log \mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]}{t_\tau}. \tag{167}$$

Substituting this form of τ into equation (164) yields

$$\mathrm{E}_s\left[g_0\left(X\right)\right] = \frac{\eta + \log \mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]}{t_\tau}. \tag{168}$$

Under the assumption that there is a one-to-one relationship between the bias, τ, and the corresponding moment generating function root, t_τ, minimizing the mean of the unbiased detector non-linearity when signal is present with respect to τ may be equivalently performed by minimization with respect to t_τ. There is also an implicit assumption that $\mathrm{E}_s\left[g_0\left(X\right)\right]$ is monotonically increasing in s so maximizing the latter is equivalent to maximizing the former. Proceeding, the partial derivative of (168) with respect to t_τ is seen to be

$$\frac{\partial}{\partial t_\tau}\mathrm{E}_s\left[g_0\left(X\right)\right] = \frac{1}{t_\tau^2}\left[t_\tau\frac{\mathrm{E}_0\left[g_0\left(X\right)\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]} - \eta - \log \mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]\right], \tag{169}$$

which is equal to zero when t_τ satisfies

$$\eta = t_\tau \frac{\mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]} - \log \mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]. \tag{170}$$

To show that this is a minimum, consider the second derivative with respect to t_τ,

$$\begin{aligned}
\frac{\partial^2}{\partial t_\tau^2}\mathrm{E}_s\left[g_0(X)\right] &= \frac{2\eta}{t_\tau^3} - \frac{1}{t_\tau^3}\left[t_\tau \frac{\mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]} - 2\log \mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]\right] \\
&\quad + \frac{t_\tau \mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]\mathrm{E}_0\left[g_0^2(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{t_\tau^2 \mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]^2} \\
&\quad - \frac{\mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]\left(\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right] + t_\tau \mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]\right)}{t_\tau^2 \mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]^2} \\
&= \frac{1}{t_\tau}\left[\frac{\mathrm{E}_0\left[g_0^2(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]} - \left(\frac{\mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]}\right)^2\right] \\
&\quad - \frac{2}{t_\tau}\frac{1}{t_\tau^2}\left[t_\tau \frac{\mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]} - \eta - \log \mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]\right] \\
&= \frac{1}{t_\tau}\left[\frac{\mathrm{E}_0\left[g_0^2(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]} - \left(\frac{\mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]}\right)^2\right] \\
&\quad - \frac{2}{t_\tau}\left(\frac{\partial}{\partial t_\tau}\mathrm{E}_s\left[g_0(X)\right]\right).
\end{aligned} \tag{171}$$

Noting that the first derivative is equal to zero when t_τ satisfies equation (170), convexity is assured at this t_τ when

$$\frac{\mathrm{E}_0\left[g_0^2(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]} - \left(\frac{\mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]}\right)^2 > 0, \tag{172}$$

or, equivalently,

$$\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]\mathrm{E}_0\left[g_0^2(X)\,\mathrm{e}^{t_\tau g_0(X)}\right] > \mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]^2. \tag{173}$$

This inequality is shown to be true by using the integral form of the Schwarz inequality, as found in Hardy [28],

$$\int_{\mathcal{E}} u^2(x)\,dx \int_{\mathcal{E}} v^2(x)\,dx \geq \left(\int_{\mathcal{E}} u(x)\,v(x)\,dx\right)^2, \tag{174}$$

where equality holds only if the non-negative functions u and v, when appropriately scaled, are equivalent[1] ,

$$Au(x) \equiv Bv(x), \tag{175}$$

where A and B are not both equal to zero. Applying the Schwarz inequality to equation (173) where

$$u(x) = \left[e^{t_\tau g_0(x)} f_X(x)\right]^{\frac{1}{2}} \tag{176}$$

and

$$v(x) = \left[g_0^2(x)\, e^{t_\tau g_0(x)} f_X(x)\right]^{\frac{1}{2}} \tag{177}$$

yields

$$\begin{aligned} \mathrm{E}\left[e^{t_\tau g_0(X)}\right] \mathrm{E}\left[g_0^2(X)\, e^{t_\tau g_0(X)}\right] &\geq \mathrm{E}\left[|g_0(X)|\, e^{t_\tau g_0(X)}\right]^2 \\ &\geq \mathrm{E}\left[g_0(X)\, e^{t_\tau g_0(X)}\right]^2, \end{aligned} \tag{178}$$

thus proving convexity and indicating that choosing t_τ to satisfy equation (170) minimizes the mean of the unbiased detector non-linearity, (168). The equality in equation (178) will hold only when

$$g_0^2(x) = c, \tag{179}$$

[1]Two functions are equivalent if they are equal except on a set of zero measure.

where c is any real constant, which is not a reasonable detector function for the block level statistics considered in this dissertation.

Now, supposing that the signal-to-interference ratio, and thus $\mathrm{E}_s\left[g_0\left(X\right)\right]$, is fixed, the bias may be chosen to maximize the asymptotic efficacy of equation (163). Consider the partial derivative of equation (163) with equation (167) with respect to t_τ

$$\frac{\partial}{\partial t_\tau}\eta \;=\; \mathrm{E}_s\left[g_0\left(X\right)\right] - \frac{\mathrm{E}_0\left[g_0\left(X\right)e^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[e^{t_\tau g_0(X)}\right]}, \tag{180}$$

which is clearly set to zero by choosing t_τ to satisfy

$$\mathrm{E}_s\left[g_0\left(X\right)\right] \;=\; \frac{\mathrm{E}_0\left[g_0\left(X\right)e^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[e^{t_\tau g_0(X)}\right]}. \tag{181}$$

That this choice maximizes the asymptotic efficacy is insured by noticing that the second derivative

$$\begin{aligned}\frac{\partial^2}{\partial t_\tau^2}\eta \;&=\; -\frac{\mathrm{E}_0\left[g_0^2\left(X\right)\mathrm{e}^{t_\tau g_0(X)}\right]\mathrm{E}_0\left[e^{t_\tau g_0(X)}\right] - \mathrm{E}_0\left[g_0\left(X\right)\mathrm{e}^{t_\tau g_0(X)}\right]^2}{\mathrm{E}_0\left[e^{t_\tau g_0(X)}\right]^2} \\ &<\; 0,\end{aligned} \tag{182}$$

which provides concavity. The strict inequality follows from equation (178) where the detector non-linearity is not allowed to have the form seen in equation (179).

The equivalence of the two methods described for choosing the moment generating function root, and thus the detector bias, may be argued using a graphically oriented proof by contradiction. The equivalence is formulated by the following theorem where $\lambda = \mathrm{E}_s\left[g_0\left(X\right)\right]$:

Theorem 4.1 If the detector bias is chosen at $\tau = \tau^*$, to minimize λ for a fixed value of the asymptotic efficacy, $\eta = \eta^*$, yielding the minimum λ^*, then τ^* also maximizes η over all valid biases, τ, for the fixed value of the unbiased detector non-linearity mean when signal is present, λ^*.

Proof: Suppose that τ^* does not maximize the asymptotic efficacy, η. Then there exists a bias, $\tilde{\tau} \neq \tau^*$, such that

$$\eta_{\tilde{\tau}}(\lambda^*) > \eta_{\tau^*}(\lambda^*), \tag{183}$$

where $\eta_\tau(\lambda)$ is the asymptotic efficacy for the detector with bias τ and non-linearity mean λ when signal is present. Using equation (163) this may be described graphically as seen in figure 3.

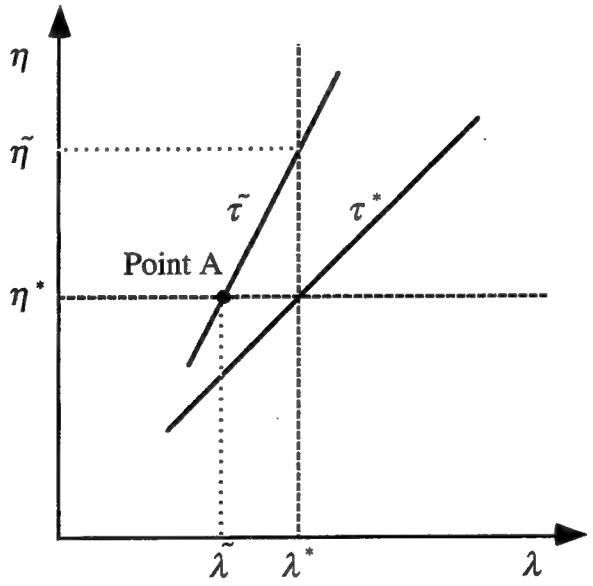

Figure 3: Efficacy versus detector non-linearity mean when signal is present.

Equation (183) results in

$$t_{\tilde{\tau}}(\lambda^* - \tilde{\tau}) \;>\; t_{\tau^*}(\lambda^* - \tau^*). \tag{184}$$

Now, consider the value of λ that sets the asymptotic efficacy to η^* for the bias $\tilde{\tau}$,

$$\begin{aligned} \eta_{\tilde{\tau}}\left(\tilde{\lambda}\right) &= t_{\tilde{\tau}}\left(\tilde{\lambda}-\tilde{\tau}\right) \\ &= \eta^* \\ &= t_{\tau^*}\left(\lambda^*-\tau^*\right). \end{aligned} \tag{185}$$

This is represented by point A in figure 3. Combining equations (184) and (185) results in the inequality

$$\begin{aligned} t_{\tilde{\tau}}\left(\lambda^*-\tilde{\tau}\right) &> t_{\tilde{\tau}}\left(\tilde{\lambda}-\tilde{\tau}\right) \\ \lambda^* &> \tilde{\lambda}, \end{aligned} \tag{186}$$

which is evident from the graph of figure 3 when $t_{\tilde{\tau}} > 0$. This result contradicts the assumption that the bias τ^* was chosen to minimize λ over all biases that yield asymptotic efficacy equal to η^*. Thus, choosing the bias to minimize the detector non-linearity mean when signal is present for a fixed asymptotic efficacy is equivalent to choosing the bias to maximize the asymptotic efficacy for a fixed detector non-linearity mean when signal is present.

4.3.2 Implementation Concerns

In practice, the asymptotic efficacy, a function of the mean time between false alarms and the mean delay to detection, would be prespecified. From the desired efficacy, equation (170) is inverted to determine the required value of t_τ. This value of the moment generating function root is used to determine the bias

required to achieve the desired efficacy by substitution into equation (167). As equations (167) and (170) are not of simple form, it is not clear if there exists a one-to-one transformation from η to t_τ or if equation (167) exists for reasonable t_τ.

From equation (170), when $t_\tau = 0$, $\eta = 0$ as well. If η is a monotone, strictly increasing function of t_τ, then for each $\eta > 0$ there exists a unique $t_\tau > 0$ that satisfies equation (170). To show that η is strictly increasing in t_τ, the partial derivative is shown to be positive,

$$\begin{aligned}
\frac{\partial \eta}{\partial t_\tau} &= \frac{\mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right] + t_\tau \mathrm{E}_0\left[g_0^2(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]} \\
&\quad - \frac{t_\tau \mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]^2}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]^2} - \frac{\mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]} \\
&= t_\tau \frac{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]\mathrm{E}_0\left[g_0^2(X)\,\mathrm{e}^{t_\tau g_0(X)}\right] - \mathrm{E}_0\left[g_0(X)\,\mathrm{e}^{t_\tau g_0(X)}\right]^2}{\mathrm{E}_0\left[\mathrm{e}^{t_\tau g_0(X)}\right]^2} \\
&> 0, \qquad (187)
\end{aligned}$$

when $t_\tau > 0$, where the results of equation (178) are used restricting $g_0(x)$ so that the strict inequality applies.

The relationship between t_τ and τ is described functionally by equation (167). Originally, t_τ was described as the value of $t > 0$ where the moment generating function of $g_\tau(X)$,

$$\begin{aligned}
M_\tau(t) &= \mathrm{E}_0\left[\mathrm{e}^{t(g_0(X)-\tau)}\right] \\
&= \mathrm{e}^{-\tau t}\mathrm{E}_0\left[\mathrm{e}^{t g_0(X)}\right], \qquad (188)
\end{aligned}$$

equals one. The fact that $t_\tau > 0$ is assured, as seen by applying Jensen's inequality, because the mean of the biased detector non-linearity is negative, (160). For a fixed value of $t > 0$, say t^*, $M_\tau(t^*)$ is a strictly decreasing function of τ, as seen by the partial derivative of equation (188),

$$\begin{aligned} \frac{\partial}{\partial \tau} M_\tau(t^*) &= -t^* e^{-t^* \tau} \mathrm{E}_0\left[e^{t^* g_0(X)}\right] \\ &< 0 \end{aligned} \tag{189}$$

for all τ. This implies that there can exist at most one value of τ, say τ^*, such that $M_{\tau^*}(t^*) = 1$. That this value exists is insured by the continuity of $M_\tau(t^*)$ in τ and the fact that

$$\lim_{\tau \to \infty} M_\tau(t^*) = 0 \tag{190}$$

and

$$\lim_{\tau \to -\infty} M_\tau(t^*) = \infty \tag{191}$$

for $t^* > 0$. This, however, does not insure that τ^* is positive.

4.3.3 Example

To illustrate the use of equations (170), (181), and (167) to determine the moment generating function root and the associated optimal bias, the GLR block level test statistic is evaluated using the GLR locally optimal detector non-linearity of equation (147). Curves relating the optimal bias to the desired

efficacy are shown in figure 4 for the adaptive dimension, $N = 1$, and various block sizes. The SIR required to achieve a specified asymptotic efficacy is shown in figure 5 as a function of the applied bias. Evaluation of equations (170) and (167) results in the optimal choice of bias which is seen to require the minimum SIR. The efficacy achieved for various SIR values as a function of the applied bias is shown in figure 6 where it is seen that the optimal choice of bias results in the maximum efficacy.

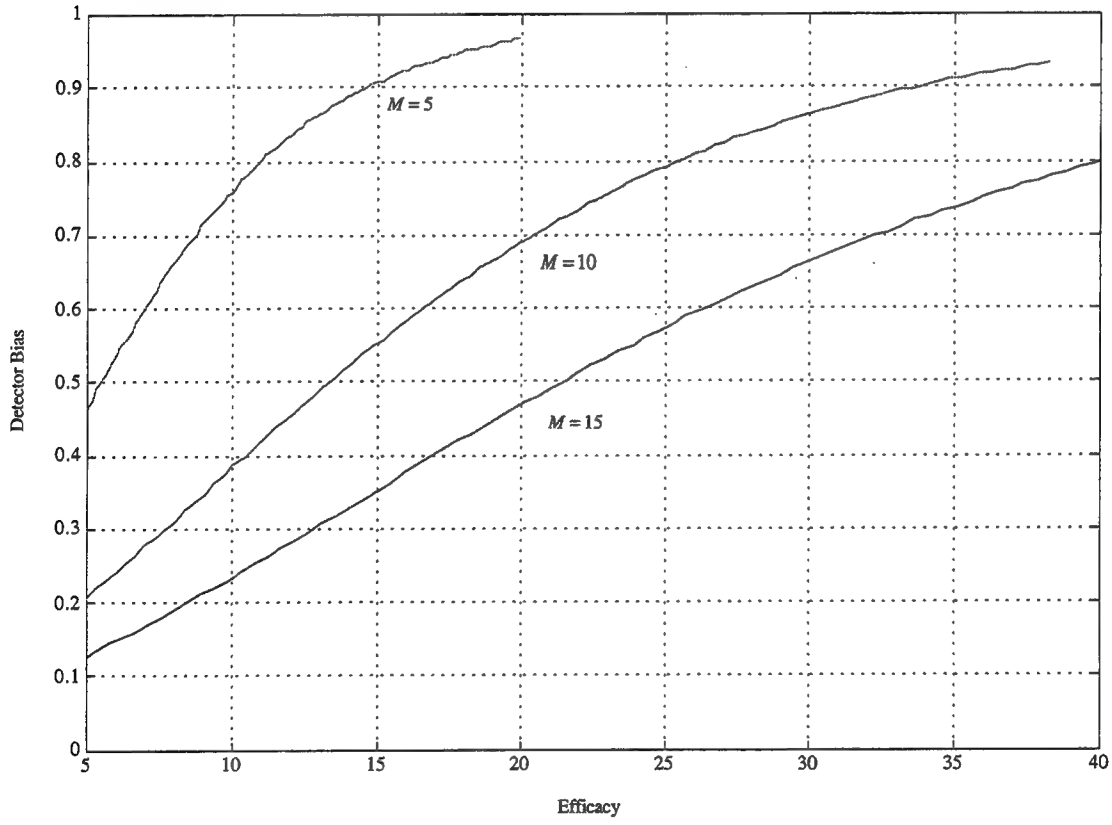

Figure 4: Optimal bias as a function of the asymptotic efficacy for $N = 1$ and block sizes $M = 5$, 10, and 15 for the GLR block level statistic with the locally optimal non-linearity.

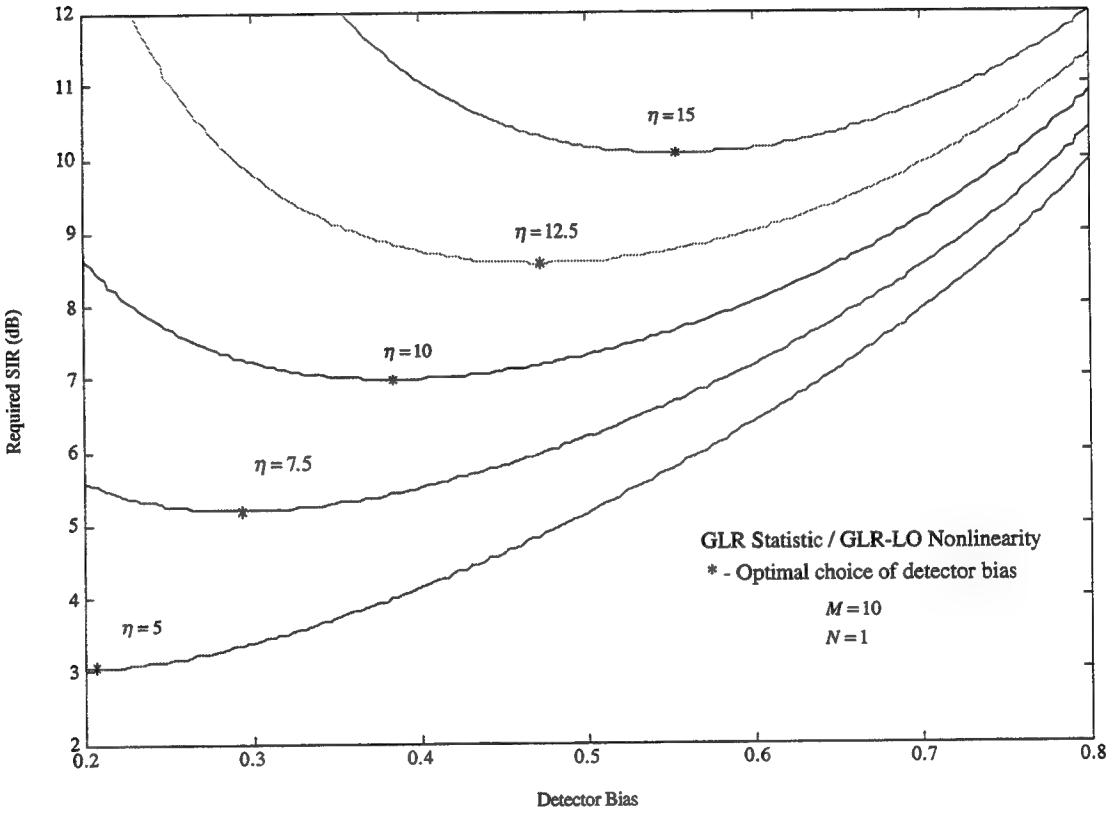

Figure 5: SIR required to achieve a specified efficacy as a function of the applied bias for $N = 1$ and $M = 10$ for the GLR block level statistic with the locally optimal non-linearity.

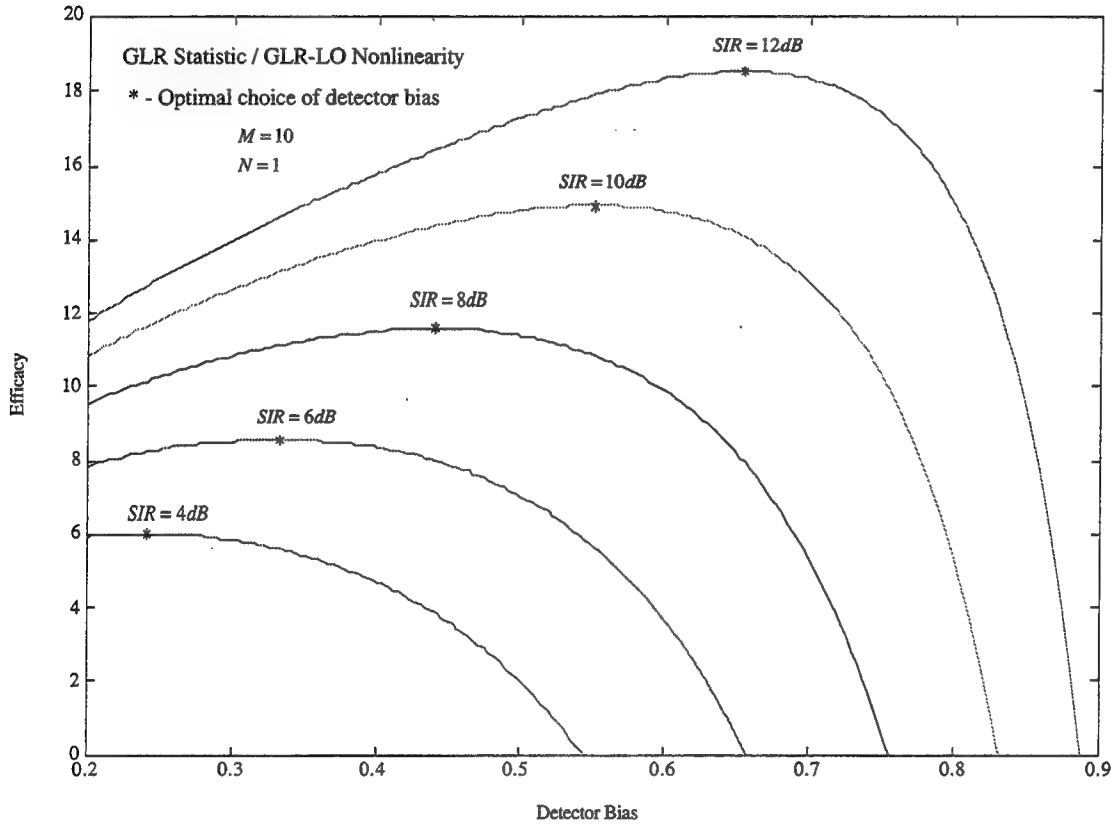

Figure 6: Efficacy achieved for specified values of signal-to-interference ratio as a function of the applied bias for $N = 1$ and $M = 10$ for the GLR block level statistic with the locally optimal non-linearity.

Chapter 5

Performance Analysis of Block Detection

In chapter 3, the GLR and AMF block level statistics were derived. In chapter 4, the log-likelihood ratio detector non-linearity was described and the locally optimal detector non-linearity was derived for the GLR and AMF block level statistics. In this chapter the combinations of the GLR and AMF block level statistics and the design SIR and locally optimal detector non-linearities are compared in terms of the asymptotic efficacy achieved for various signal-to-interference ratios. The best of the four combinations will then be examined as a function of the block size and the adaptive dimension which is the number of elements in each data vector.

5.1 Preliminaries

Adaptive algorithms, including the one presented in this dissertation, often require the estimation and inversion of covariance matrices. When the dimension

of the data is large, for instance an array with a large number of sensors, adequate estimation of the covariance matrix requires a large number of samples and the inversion of the covariance matrix estimate becomes numerically inhibitive as well as potentially unstable. Matrix preprocessing of the array data to a smaller dimension reduces numerical computations often with near equivalent average performance as shown by Owsley and Abraham [29] and also provides improved statistical performance as shown by Burgess and VanVeen [30]. Matrix preprocessing entails a linear transformation of P dimensional sensor data to N dimensional data using a constant N-by-P matrix. Gray's [31] beamspace preprocessing algorithm is of particular interest because the adaptive dimension can be dramatically reduced with minimal loss in performance. The performance may be quantified by the array gain improvement (AGI), as defined by Owsley [32], of the preprocessor as a function of the adaptive dimension. The signal-to-interference ratio for the preprocessed data is the array gain improvement multiplied by the SIR for the conventional processing case, which, for beam space preprocessing, has the form

$$SIR(N) = AGI(N)\, SIR(N = 1)\,. \tag{192}$$

As the adaptive dimension, N, is increased towards the total number of sensors, the AGI increases. In certain types of interference environments, in particular those with strong plane wave interferers, the array gain improvement can be quite large, for instance, on the order of ten decibels.

The efficacy, as described in chapter 2, is a function of the average number of samples between false alarms and the average delay before detection. By implementing Page's test on blocks of data, these parameters are defined at the block level. However, the detection algorithm must be designed based on the performance at the sample level. If M is the block size, if η_B is the asymptotic efficacy at the block level, and if T_B and D_B are respectively the average number of blocks between false alarms and before detection, the sample level asymptotic efficacy may be described as

$$\begin{aligned} \eta &= \frac{\log(MT_B)}{MD_B} \\ &= \frac{\log(T_B)}{MD_B} + \frac{\log(M)}{MD_B} \\ &= \frac{\eta_B}{M} + \frac{\log(M)}{MD_B}. \end{aligned} \tag{193}$$

Varying D_B results in upper and lower bounds on the sample level asymptotic efficacy in terms of the block size and block level efficacy

$$\frac{\eta_B}{M} < \eta \leq \frac{\eta_B}{M} + \frac{\log(M)}{M}. \tag{194}$$

Equation (194) may be solved for the block level efficacy yielding the form

$$M\eta - \log(M) \leq \eta_B < M\eta. \tag{195}$$

The following performance analysis describes the performance of the block detection algorithm in terms of either the signal-to-interference ratio required to

achieve a specified efficacy or one of the sample level efficacy bounds of equation (194) achieved by a specified SIR. Evaluation of the efficacy requires knowledge of the mean of the detector non-linearity as a function of the signal-to-interference ratio and the moment generating function root of the detector non-linearity. Both of these require the probability density function of the block level statistic. Numerical integration is used to form tables of the PDF of the block level statistics when no closed form exists. Similarly, tables are formed for the design SIR non-linearities and the locally optimal non-linearity for the AMF statistic.

5.2 Performance Analysis

The first objective is to evaluate the performance of all the combinations of the block level statistics and the detector non-linearities. Figures 7 and 8 respectively contain curves of the lower bound on the sample level efficacy for the design SIR and locally optimal non-linearities comparing the GLR and AMF statistics as a function of the signal-to-interference ratio for the block size $M = 15$ and the adaptive dimensions $N = 1$, 5, and 10. The design SIR non-linearities were chosen to be optimal for a signal-to-interference ratio of $s = 0$ dB and the locally optimal non-linearities were implemented with a bias chosen to maximize the efficacy for $s = 0$ dB. Note that for the $N = 1$ case, the GLR and AMF block level statistics are identical. It is clear that for both non-linearities the GLR statistic outperforms the AMF statistic.

Figures 9 and 10 respectively contain curves of the lower bound on the sample level the efficacy for the AMF and GLR statistics comparing the design SIR and locally optimal non-linearities as a function of the signal-to-interference ratio for the block size $M = 15$ and the adaptive dimensions $N = 1$, 5, and 10. Both figures reflect equivalent performance at $s = 0$ dB, a result of the non-linearities being optimized by design or bias for this level. The locally optimal non-linearity performs worse than the design SIR non-linearity for the AMF statistic and yet performs slightly better than the design SIR non-linearity for the GLR statistic. The degradation increases as the SIR is displaced from the design value, especially for smaller values of SIR. The difference in performance at displaced SIR values decreases as the adaptive dimension increases for the GLR statistic.

These comparisons indicate that the locally optimal detector non-linearity operating on the GLR statistic provides the best performance while the locally optimal detector non-linearity operating on the AMF statistic provides the worst performance. This is an appealing result because the GLR statistic has a simpler conditional probability density function, particularly when no signal is present, as the PDF is central chi-squared. Additionally, the locally optimal non-linearity for the GLR statistic has a closed form representation where the design SIR non-linearities and the locally optimal non-linearity for the AMF statistic require a quantized implementation. All further analysis will be performed for the GLR statistic with its locally optimal non-linearity.

It is now desired to analyze the performance of the detector as a function of the adaptive dimension. Figure 11 contains curves of the SIR required to achieve a particular value of efficacy as a function of the adaptive dimension. The SIR is chosen so that the block level efficacy meets the upper bound of equation (195). The bias chosen to implement the detector minimizes the SIR required to achieve the desired efficacy. Efficacies ranging from 0.1 to 0.9 are shown for a block size of $M = 15$ and adaptive dimensions varying from $N = 1$ to $N = 13$. The utility of these curves is shown by the dashed curve superimposed on the constant efficacy contours. This curve is the signal-to-interference ratio of equation (192) for beamspace preprocessed array data as a function of the adaptive dimension of the preprocessor for a specific interference scenario and beamforming look direction. Varying the SIR at the output of a conventional beamformer, $SIR(N = 1)$, simply shifts the dashed curve up or down which in turn indicates the adaptive dimension that provides the best performance. These curves may be used to determine the preprocessor type and dimension that provide the best performance in a specific interference scenario. The scenario for this example consisted of a thirty-two sensor, half wavelength, equally spaced line array beamformed broadside to the array. The interferences consisted of plane waves at -2, 2, and 5 degrees from broadside respectively at the levels 13, 10, and 10 dB above spatially uncorrelated noise and a Butterworth angularly extended interfering signal at 4 degrees and 10 dB with a spreading coefficient of $\rho = 0.5$. If the interference scenario was restricted to spatially uncorrelated

noise, the SIR curve would be constant as the adaptive dimension increased which would indicate that conventional beamforming should be used to maximize the efficacy. This result is expected as the conventional beamformer maximizes the beam output SIR for spatially uncorrelated background noise.

In the previous analyses, the block size has been held fixed at $M = 15$. The block size may be limited from above by the degree of the non-stationarity of the interfering signals as well as the loss in time resolution and is bounded below by $M > N + 1$. The signal-to-interference ratios required to achieve the upper and lower bounds on the block level efficacy of equation (195) for an asymptotic efficacy of $\eta = 0.5$ for $N = 1$, 5, and 10 are shown in figure 12 as a function of the block size. The lower bound curve for the $N = 1$ case clearly indicates that a block size of $M = 5$ yields the minimum required SIR to achieve $\eta = 0.5$. These curves may be used to choose a block size that provides adequate performance without violating stationarity assumptions or substantial loss in time resolution.

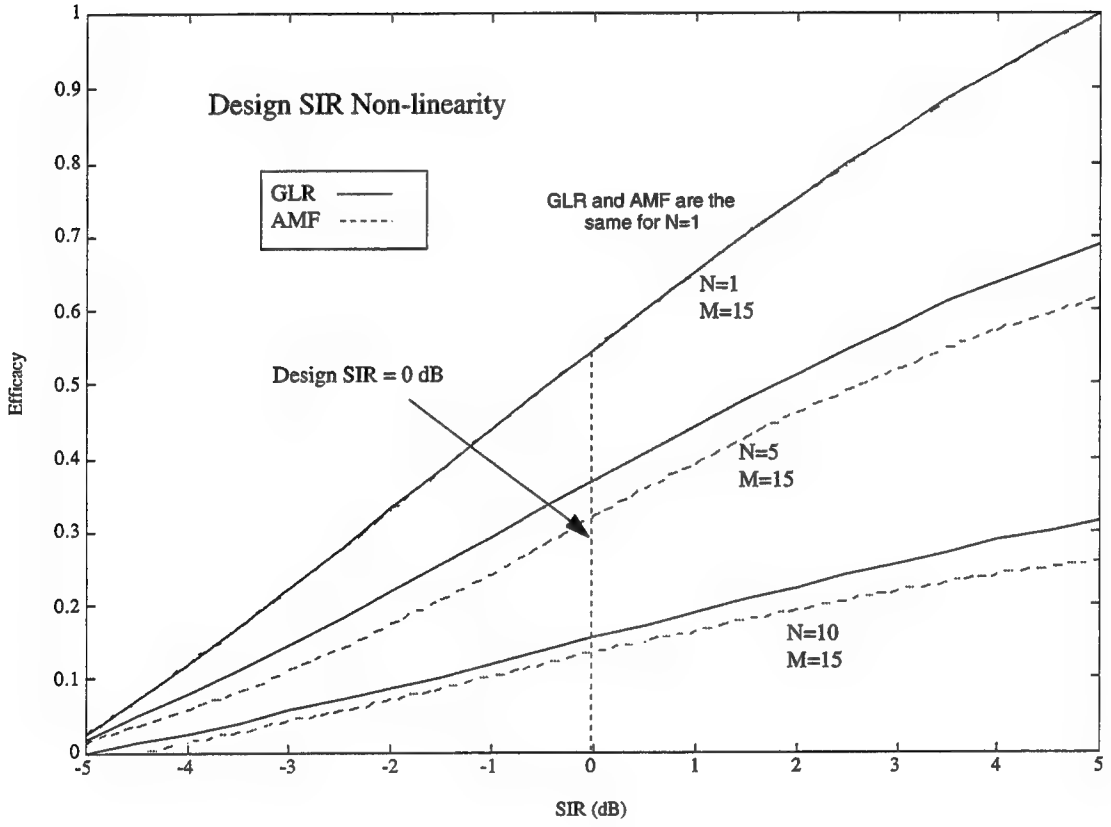

Figure 7: Efficacy curves for the design SIR non-linearity for the GLR and AMF statistics for $M = 15$, $N = 1$, 5, and 10.

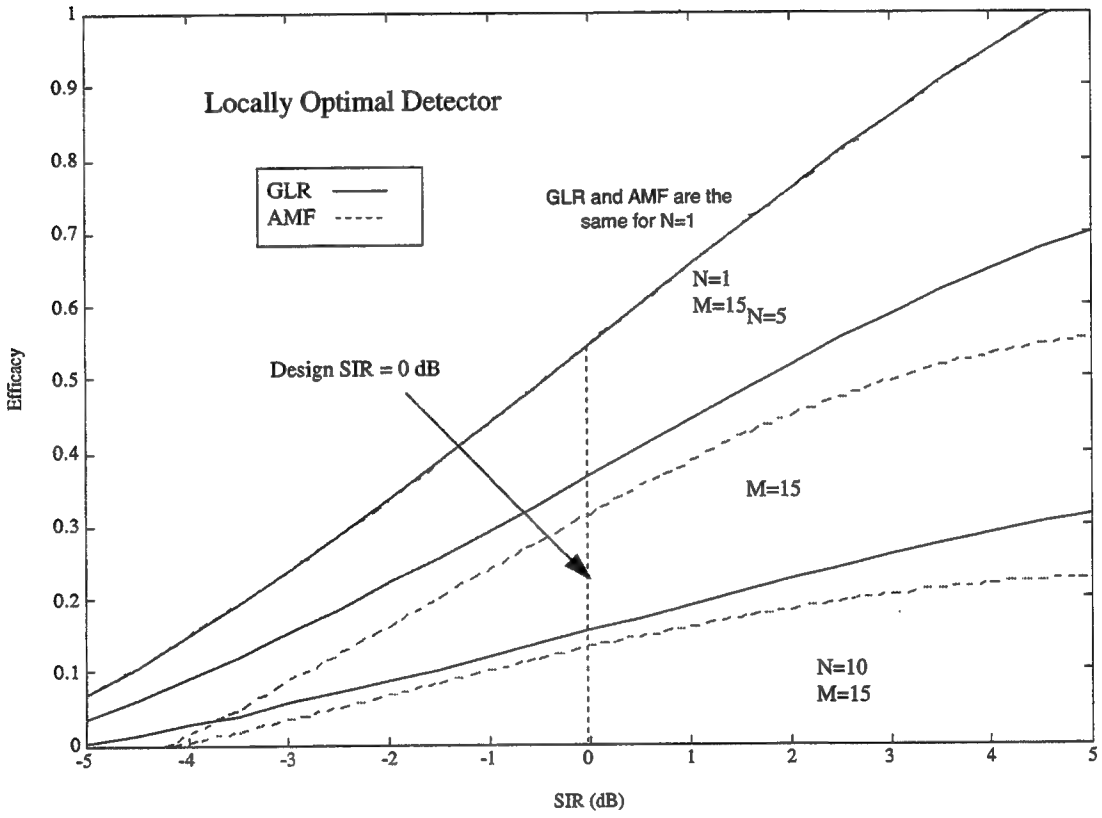

Figure 8: Efficacy curves for the locally optimal non-linearity for the GLR and AMF statistics for $M = 15$, $N = 1$, 5, and 10.

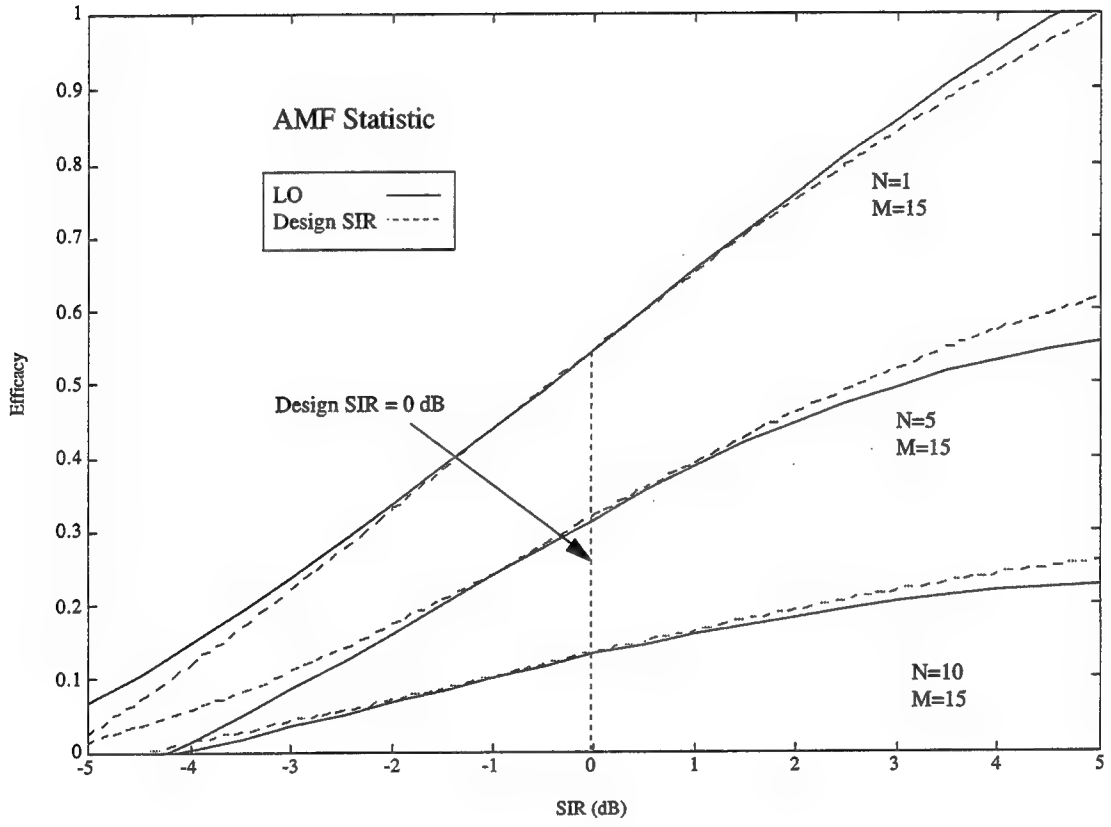

Figure 9: Efficacy curves for the AMF statistic for the design SIR and locally optimal non-linearities for $M = 15$, $N = 1$, 5, and 10.

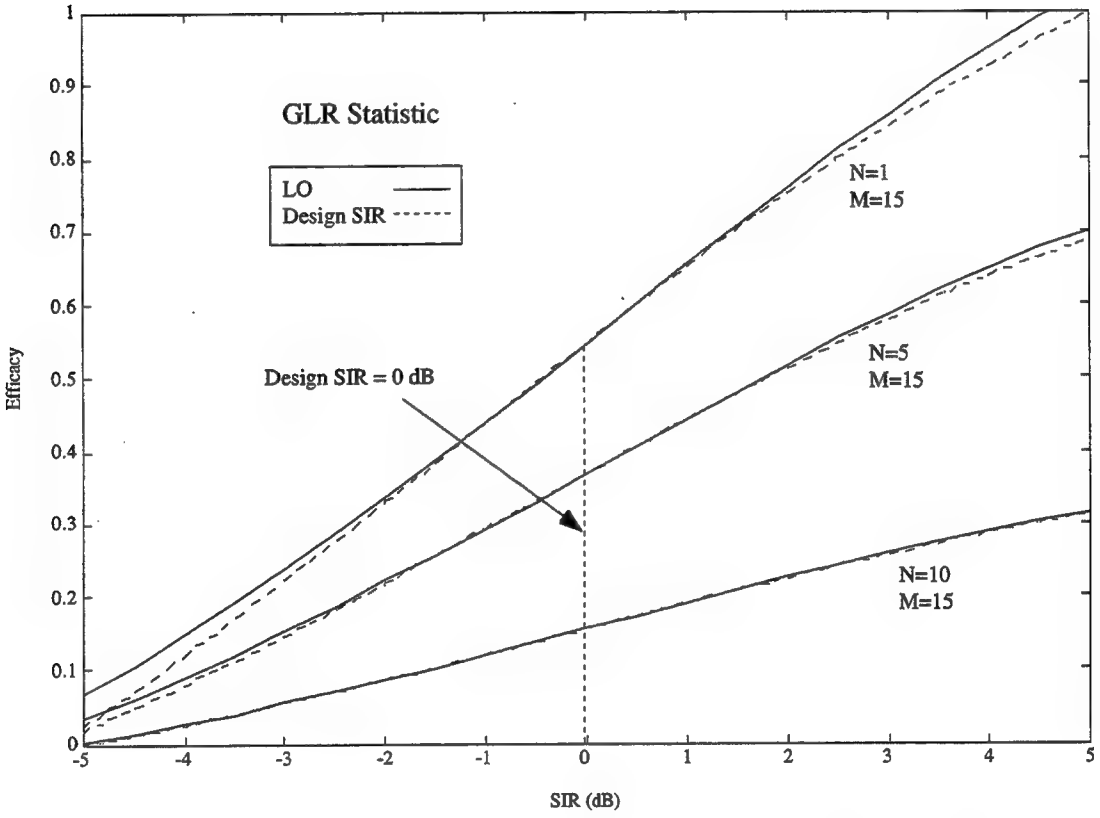

Figure 10: Efficacy curves for the GLR statistic for the design SIR and locally optimal non-linearities for $M = 15$, $N = 1$, 5, and 10.

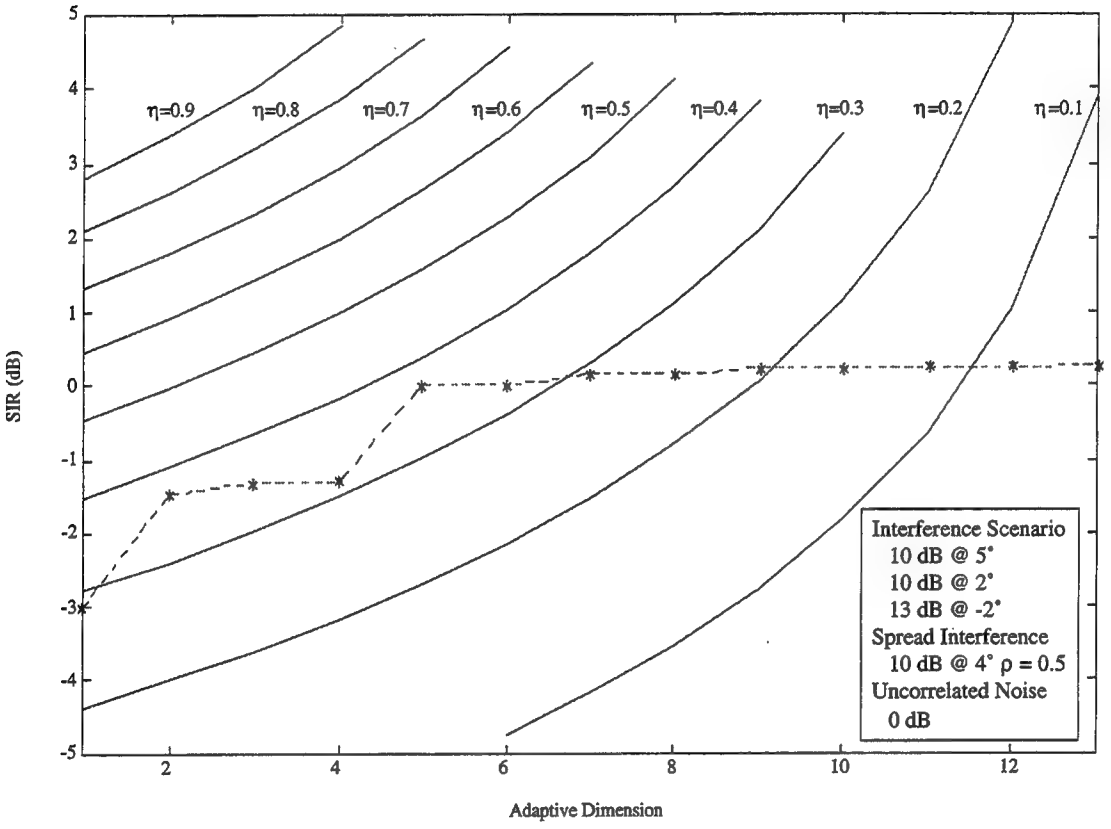

Figure 11: SIR required to achieve constant efficacy as a function of the adaptive dimension for the GLR locally optimal detector with a block size of $M = 15$. The dashed line represents the SIR for a broadside look direction for a specific interference scenario with beamspace preprocessing of a sixty-four sensor, half wavelength, equally spaced line array where the conventional beam output SIR is $SIR(N = 1) = -3$ dB.

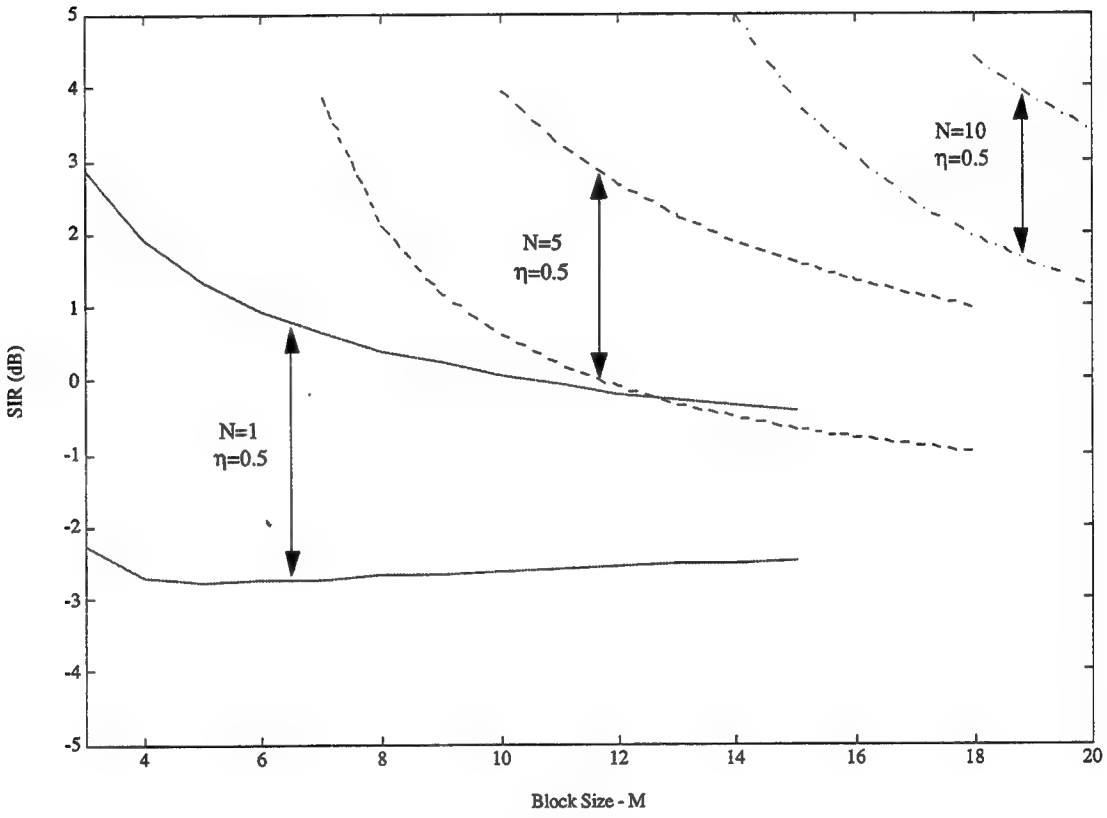

Figure 12: Upper and lower bounds on the SIR required to achieve a sample level efficacy of $\eta = 0.5$ as a function of block size for $N = 1$, 5, and 10 for the GLR locally optimal detector.

Chapter 6

Post Block Processing

Once Page's test, implemented using the detector non-linearities described in chapter 4, has indicated that a signal has been detected, further processing of the data may provide improved time resolution of the estimate of the onset time. In speech processing or classification applications, this will provide improved segmentation of short time duration signals. Page's test operates on the block level data, providing a coarse estimate of the onset time of the signal. Assuming that the interference is stationary over a length of time that straddles the signal onset time, a maximum likelihood estimate of the signal onset time may be found for the deterministic signal model described in section 3.1.1.

6.1 Derivation

Let the block of data to be used for the estimation of the signal onset time be the N-dimensional vectors

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M, \tag{196}$$

where M is the size of the block. The block size should be as large as possible without violating the stationarity of the interference. If the signal, with unknown complex amplitude θ, begins at sample P, the data will be distributed as,

$$\mathbf{x}_i \sim \begin{cases} \mathcal{CN}_N(\mathbf{0}, \boldsymbol{\Sigma}) & i < P \\ \mathcal{CN}_N(\theta\mathbf{d}, \boldsymbol{\Sigma}) & i \geq P \end{cases}, \tag{197}$$

under the deterministic signal model. The joint probability density function of all the data in the block is

$$\begin{aligned} f(\mathbf{X}|\boldsymbol{\Sigma}, \theta, P) &= \prod_{i=1}^{P-1} \frac{\exp\left(-\mathbf{x}_i^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_i\right)}{\pi^N |\boldsymbol{\Sigma}|} \prod_{i=P}^{M} \frac{\exp\left[-(\mathbf{x}_i - \theta\mathbf{d})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \theta\mathbf{d})\right]}{\pi^N |\boldsymbol{\Sigma}|} \\ &= \operatorname{etr}\left\{ -\boldsymbol{\Sigma}^{-1} \begin{bmatrix} \sum_{i=1}^{M} \mathbf{x}_i \mathbf{x}_i^H + (M - P + 1)\,\theta\theta^* \mathbf{d}\mathbf{d}^H \\ -\theta\mathbf{d} \sum_{i=P}^{M} \mathbf{x}_i^H - \theta^* \sum_{i=P}^{M} \mathbf{x}_i \mathbf{d}^H \end{bmatrix} \right\} \\ &\quad \times \frac{1}{\pi^{MN} |\boldsymbol{\Sigma}|^M} \\ &= \left[\frac{1}{\pi^N |\boldsymbol{\Sigma}|} \operatorname{etr}\left\{ -\boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbf{S} - \beta_P \theta \mathbf{d} \tilde{\mathbf{x}}_P^H \\ -\beta_P \theta^* \tilde{\mathbf{x}}_P \mathbf{d}^H + \beta_P \theta\theta^* \mathbf{d}\mathbf{d}^H \end{bmatrix} \right\} \right]^M \\ &= \left[\frac{1}{\pi^N |\boldsymbol{\Sigma}|} \operatorname{etr}\left\{ -\boldsymbol{\Sigma}^{-1} \mathbf{B}_P(\theta) \right\} \right]^M, \end{aligned} \tag{198}$$

where

$$\mathbf{X} = \left[\mathbf{x}_1 \;\; \mathbf{x}_2 \;\; \cdots \;\; \mathbf{x}_M \right], \tag{199}$$

$$\mathbf{S} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{x}_i \mathbf{x}_i^H, \tag{200}$$

$$\beta_P = \frac{M-P+1}{M}, \tag{201}$$

$$\tilde{\mathbf{x}}_P = \frac{1}{M-P+1} \sum_{i=P}^{M} \mathbf{x}_i, \tag{202}$$

and

$$\mathbf{B}_P(\theta) = \mathbf{S} - \beta_P \theta \mathbf{d} \tilde{\mathbf{x}}_P^H - \beta_P \theta^* \tilde{\mathbf{x}}_P \mathbf{d}^H + \beta_P \theta \theta^* \mathbf{d}\mathbf{d}^H. \tag{203}$$

The maximum likelihood estimate for the unknown start time requires the maximization of equation (198) with respect to $\mathbf{\Sigma}$, θ, and P. It is recognized, as seen in section 3.1.1, that the maximization over $\mathbf{\Sigma}$ requires choosing

$$\hat{\mathbf{\Sigma}} = \mathbf{B}_P(\theta). \tag{204}$$

The resulting likelihood function is

$$f\left(\mathbf{X}|\hat{\mathbf{\Sigma}}, \theta, P\right) = \left[\frac{\mathrm{e}^{-N}}{\pi^N \left|\mathbf{B}_P(\theta)\right|} \right]^M, \tag{205}$$

which may be equivalently maximized over θ by minimizing the determinant of $\mathbf{B}_P(\theta)$. The closed form of the determinant is seen to be

$$\begin{aligned} \left|\mathbf{B}_P(\theta)\right| &= \left|\mathbf{S} - \beta_P \theta \mathbf{d} \tilde{\mathbf{x}}_P^H - \beta_P \theta^* \tilde{\mathbf{x}}_P \mathbf{d}^H + \beta_P \theta \theta^* \mathbf{d}\mathbf{d}^H\right| \\ &= |\mathbf{S}| \left|\mathbf{I}_N + \beta_P \mathbf{S}^{-1} \mathbf{D}\mathbf{E}\mathbf{D}^H\right|, \end{aligned} \tag{206}$$

where

$$\mathbf{D} = \left[\theta \mathbf{d} \;\; \tilde{\mathbf{x}}_P \right], \tag{207}$$

and

$$\mathbf{E} = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}. \tag{208}$$

Using the determinant identity

$$|\mathbf{I}_n + \mathbf{AB}| = |\mathbf{I}_m + \mathbf{BA}|, \tag{209}$$

where $\mathbf{A}$ is n-by-m and $\mathbf{B}$ is m-by-n, as found in Kailath [33], results in

$$\begin{aligned}
|\mathbf{B}_P(\theta)| &= |\mathbf{S}| \left|\mathbf{I}_N + \beta_P \left(\mathbf{S}^{-1}\mathbf{DE}\right)\mathbf{D}^H\right| \\
&= |\mathbf{S}| \left|\mathbf{I}_2 + \beta_P \mathbf{D}^H\mathbf{S}^{-1}\mathbf{DE}\right| \qquad (210) \\
&= |\mathbf{S}| \left|\mathbf{I}_2 + \beta_P \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}\right| \\
&= |\mathbf{S}| \begin{vmatrix} 1+\beta_P(d_{11}-d_{12}) & -\beta_P d_{11} \\ \beta_P(d_{21}-d_{22}) & 1-\beta_P d_{21} \end{vmatrix} \\
&= |\mathbf{S}| \left[(1+\beta_P(d_{11}-d_{12}))(1-\beta_P d_{21}) + \beta_P^2 d_{11}(d_{21}-d_{22})\right] \\
&= |\mathbf{S}| \left[1+\beta_P d_{11}(1-\beta_P d_{22}) - \beta_P(d_{21}+d_{12}) + \beta_P^2 d_{21}d_{12}\right], \qquad (211)
\end{aligned}$$

where $\mathbf{I}_2$ is the two dimensional identity matrix and

$$\mathbf{D}^H\mathbf{S}^{-1}\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}. \tag{212}$$

Substituting equation (207) into (212) results in

$$d_{11} = \theta\theta^* \mathbf{d}^H\mathbf{S}^{-1}\mathbf{d} \tag{213}$$

$$d_{12} = \theta^* \mathbf{d}^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P \tag{214}$$

$$d_{21} = \theta \tilde{\mathbf{x}}_P^H \mathbf{S}^{-1} \mathbf{d} \tag{215}$$

$$d_{22} = \tilde{\mathbf{x}}_P^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P, \tag{216}$$

which, when substituted into equation (211) and completing the complex square for θ, results in

$$\begin{aligned} |\mathbf{B}_P(\theta)| &= |\mathbf{S}| \left[1 + \beta_P \left(G\theta\theta^* - \theta \tilde{\mathbf{x}}_P^H \mathbf{S}^{-1} \mathbf{d} - \theta^* \mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P\right)\right] \\ &= |\mathbf{S}| \left[1 - \beta_P \frac{\left|\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P\right|^2}{G} + \beta_P G \begin{pmatrix} \theta\theta^* - \theta \frac{\tilde{\mathbf{x}}_P^H \mathbf{S}^{-1} \mathbf{d}}{G} \\ -\theta^* \frac{\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P}{G} + \left|\frac{\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}}{G}\right|^2 \end{pmatrix}\right] \\ &= |\mathbf{S}| \left[1 - \beta_P \frac{\left|\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P\right|^2}{G} + \beta_P G \left|\theta - \frac{\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P}{G}\right|^2\right], \end{aligned} \tag{217}$$

where

$$G = \mathbf{d}^H \mathbf{S}^{-1} \mathbf{d} \left(1 - \beta_P \tilde{\mathbf{x}}_P^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P\right) + \beta_P \left|\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P\right|^2. \tag{218}$$

If $G > 0$, equation (217) is minimized by setting

$$\hat{\theta} = \frac{\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P}{G}, \tag{219}$$

which results in

$$\begin{aligned} f\left(\mathbf{X}|\hat{\boldsymbol{\Sigma}}, \hat{\theta}, P\right) &= \left[\frac{(\pi \mathrm{e})^{-N}}{|\mathbf{S}| \left(1 - \beta_P \frac{|\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P|^2}{G}\right)}\right]^M \\ &= \frac{(\pi \mathrm{e})^{-NM}}{|\mathbf{S}|^M} \left[1 + \frac{\beta_P \left|\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P\right|^2}{\mathbf{d}^H \mathbf{S}^{-1} \mathbf{d} \left(1 - \beta_P \tilde{\mathbf{x}}_P^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P\right)}\right]^M. \end{aligned} \tag{220}$$

The maximum likelihood estimate of P is found by choosing the value that maximizes equation (220). This requires a search over values of P that do not violate

the $G > 0$ condition nor the requirement that the estimate of the interference covariance matrix, $\hat{\mathbf{\Sigma}}$, be positive definite. Considering the first requirement,

$$\begin{aligned} 0 &< G \\ 0 &< \mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}\left(1-\beta_P\tilde{\mathbf{x}}_P^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P\right)+\beta_P\left|\mathbf{d}^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P\right|^2 \\ 1 &< \beta_P\left(\tilde{\mathbf{x}}_P^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P-\frac{\left|\mathbf{d}^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P\right|^2}{\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}}\right) \\ P &> M\left(1-\frac{\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}}{\left(\mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}\right)\left(\tilde{\mathbf{x}}_P^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P\right)-\left|\mathbf{d}^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P\right|^2}\right)+1, \end{aligned} \tag{221}$$

it is seen that a lower bound is placed on P. The bound itself is a function of P, thus, it must be evaluated for all potential start times from 1 to M.

The second requirement was that the estimate of the interference covariance matrix, $\mathbf{B}_P(\theta)$, be positive definite. This is shown by first proving that $\mathbf{B}_P\left(\hat{\theta}\right)$ is positive semi-definite for all θ using a quadratic form and then by requiring the determinant of $\mathbf{B}_P\left(\hat{\theta}\right)$ to be greater than zero. By placing the block data columnwise into the N-by-M matrix $\mathbf{X}$, the matrix $\mathbf{S}$ and the vector $\tilde{\mathbf{x}}_P$ may be written as

$$\mathbf{S}=\frac{1}{M}\mathbf{X}\mathbf{X}^H \tag{222}$$

and

$$\tilde{\mathbf{x}}_P=\frac{1}{M-P+1}\mathbf{X}\mathbf{e}_P, \tag{223}$$

where $\mathbf{e}_P$ represents an M-by-1 dimensional vector with zeros in the first $P-1$ elements and ones elsewhere,

$$\mathbf{e}_P = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{224}$$

The matrix $\mathbf{B}_P(\theta)$ may now be written in the form

$$\begin{aligned} \mathbf{B}_P(\theta) &= \mathbf{S} - \beta_P \tilde{\mathbf{x}}_P \tilde{\mathbf{x}}_P^H + \beta_P \begin{bmatrix} \tilde{\mathbf{x}}_P & -\theta\mathbf{d} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_P & -\theta\mathbf{d} \end{bmatrix}^H \\ &= \frac{1}{M}\mathbf{X}\mathbf{X}^H - \frac{\beta_P}{(M-P+1)^2}\mathbf{X}\mathbf{e}_P\mathbf{e}_P^H\mathbf{X}^H + \beta_P\mathbf{V}\mathbf{V}^H \\ &= \frac{1}{M}\mathbf{X}\left[\mathbf{I}_M - \frac{\mathbf{e}_P\mathbf{e}_P^H}{M-P+1}\right]\mathbf{X}^H + \beta_P\mathbf{V}\mathbf{V}^H, \end{aligned} \tag{225}$$

where $\mathbf{I}_M$ is the M dimensional identity matrix and

$$\mathbf{V} = \begin{bmatrix} \tilde{\mathbf{x}}_P & -\theta\mathbf{d} \end{bmatrix}. \tag{226}$$

Applying this form of $\mathbf{B}_P(\theta)$ to a quadratic form in $\mathbf{y}$ results in

$$\mathbf{y}^H\mathbf{B}_P(\theta)\mathbf{y} = \frac{1}{M}\mathbf{y}^H\mathbf{X}\mathbf{U}\mathbf{X}^H\mathbf{y} + \beta_P\mathbf{y}^H\mathbf{V}\mathbf{V}^H\mathbf{y}, \tag{227}$$

where

$$\mathbf{U} = \left[\mathbf{I}_M - \frac{\mathbf{e}_P\mathbf{e}_P^H}{M-P+1}\right]. \tag{228}$$

As the latter term in equation (227) is clearly the inner product of the vector $\mathbf{V}^H\mathbf{y}$, it must be greater than or equal to zero, where the equality may occur because $\mathbf{V}$ is at most rank 2. The matrix $\mathbf{X}$ has rank equal to N because its product, $\mathbf{S} = \frac{1}{M}\mathbf{X}\mathbf{X}^H$, has rank equal to N. The matrix $\mathbf{U}$ is recognized as a projection matrix with rank equal to $M-1$. Thus, any quadratic form involving the matrix $\mathbf{U}$ must lie between zero and one,

$$0 \leq \mathbf{z}^H\mathbf{U}\mathbf{z} \leq 1. \tag{229}$$

The first term of equation (227) is a quadratic form of the matrix $\mathbf{U}$ and the M dimensional vector $\mathbf{X}^H\mathbf{y}$. Thus, the quadratic form

$$\mathbf{y}^H\mathbf{B}_P(\theta)\mathbf{y} \geq 0, \tag{230}$$

and the matrix $\mathbf{B}_P(\theta)$ is positive semi-definite for all θ.

The necessary requirement that the determinant of $\mathbf{B}_P\left(\hat{\theta}\right)$ be positive yields the constraint,

$$\begin{aligned}
0 &< \left|\mathbf{B}_P\left(\hat{\theta}\right)\right| \\
0 &< |\mathbf{S}|\left[1-\beta_P\frac{\left|\mathbf{d}^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P\right|^2}{G}\right] \\
0 &< G-\beta_P\left|\mathbf{d}^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P\right|^2 \\
0 &< \mathbf{d}^H\mathbf{S}^{-1}\mathbf{d}\left(1-\beta_P\tilde{\mathbf{x}}_P^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P\right) \\
P &> M\left(1-\frac{1}{\tilde{\mathbf{x}}_P^H\mathbf{S}^{-1}\tilde{\mathbf{x}}_P}\right)+1,
\end{aligned} \tag{231}$$

which places a tighter bound on P than equation (221). The resulting maximum likelihood estimator of P may be written as

$$\hat{P} = \arg\max_{P \in \Phi} \frac{(M-P+1)\left|\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P\right|^2}{M-(M-P+1)\,\tilde{\mathbf{x}}_P^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P}, \tag{232}$$

where

$$\Phi = \left\{ P: \;\; 1 \le P \le M \;\; \text{and} \;\; P > M\left(1 - \frac{1}{\tilde{\mathbf{x}}_P^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P}\right) + 1 \right\}. \tag{233}$$

6.2 Performance Analysis

The post block processing algorithm of equations (232) and (233) is examined in this section for performance as a function of the true start time, the signal-to-interference ratio, and the adaptive dimension. Two-hundred trials were run for the adaptive dimensions $N = 1$, 5, and 10, and for block sizes $M = 10$, 15, 20, 25, and 30, excepting the combination $N = 10$ and $M = 10$. Signal-to-interference ratios ranging from $s = -5$ dB to 10 dB in one decibel increments were considered. For each case, the estimated start time was computed for each possible true start time. The results for the cases were very similar, exhibiting expected trends. Thus, the $N = 5$, $M = 20$ case is used to present these trends. Histograms for several true start times for the $s = 5$ dB case are found in figure 13. Note that when the signal starts in the latter portion of the block (the upper histogram), there is a skewness to the left that is larger than the corresponding skewness to the right displayed by the histogram of the signal that starts in the initial portion of the block (the lower histogram). This may be explained by there being very

few samples with signal present when the signal starts in the latter portion of the block, causing a more difficult estimation problem. Figure 14 contains histograms for a fixed true start time for signal-to-interference ratios varying from -5 to 10 dB in three decibel increments. The mean squared error is estimated from the histograms and shown in figure 15 as a function of the true start time for SIR values of -5, 0, 5, and 10 dB. The asymmetry of the error noted in the histograms is also seen here where the mean squared error is largest for signals that start in the latter quarter of the block. Weak signals incur larger errors when they start near either boundary, with the best performance somewhere near the middle of the block. The mean squared error is averaged over all possible true start times and plotted against the signal-to-interference ratio for the adaptive dimensions $N = 1$, 5, and 10 in figure 16. Here it is seen that there is minimal loss between the $N = 1$ case and the $N = 5$ case indicating that the performance losses associated with processing multivariate data do not accrue rapidly.

6.3 Invariance

In considering equations (232) and (233), it is not clear if the post block maximum likelihood estimate of the signal onset time is invariant to the interference covariance matrix. By applying the whitening transformation of section 3.1.2 described by equations (43)-(55) it may be shown that the probability density function of the forms that describe the maximum likelihood estimator of the

signal onset time,

$$\mathbf{d}^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P \tag{234}$$

and

$$\tilde{\mathbf{x}}_P^H \mathbf{S}^{-1} \tilde{\mathbf{x}}_P, \tag{235}$$

depend on the interference covariance matrix only through the parameter β described by equations (47) and (48). The parameter β may be thought of as a whitened complex strength parameter with magnitude equal to the signal-to-interference ratio. The dependence of the post block processing performance on the SIR is acceptable. The effect of the dependence on the *whitened* phase of the signal is not clear, however, it is much better than dependence on an N dimensional covariance matrix. The simulations of section 6.2 were conducted with an identity matrix for an interference covariance matrix and for $\beta = \sqrt{s}$.

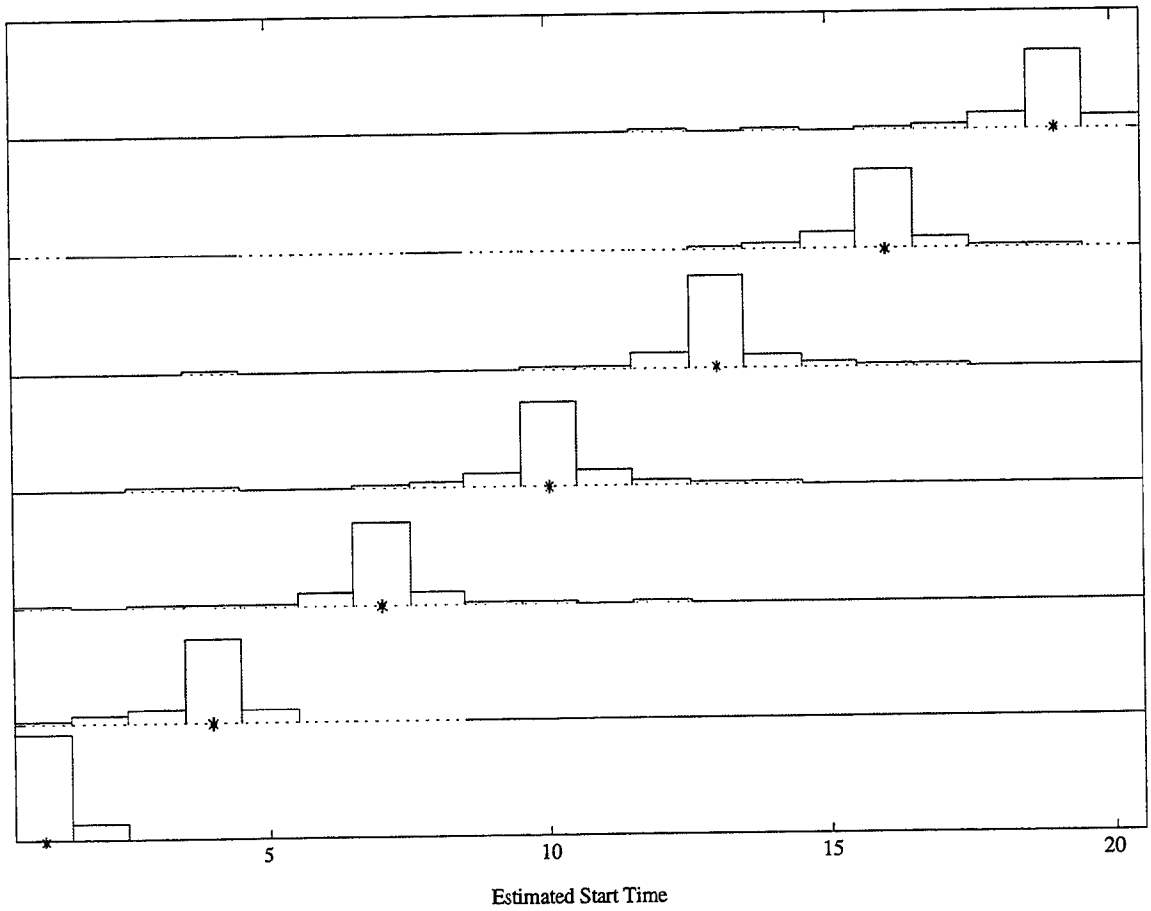

Figure 13: Histograms of the onset time estimate for several true start times for an SIR of 5 dB for $N = 5$ and $M = 20$. The asterisk represents the true start time. Each histogram has been scaled by the same value to facilitate comparison.

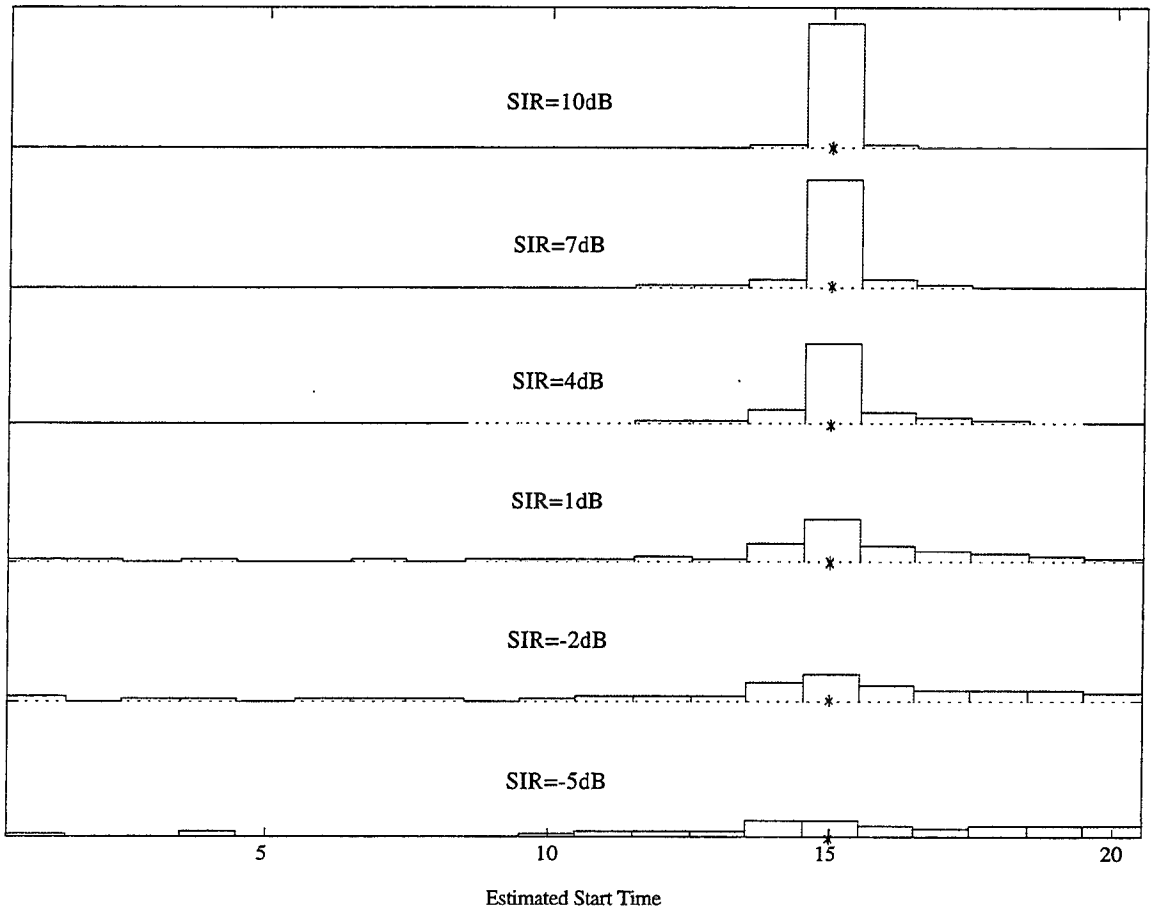

Figure 14: Histograms of the onset time estimate for several SIR values for $N = 5$ and $M = 20$. The asterisk represents the true start time. Each histogram has been scaled by the same value to facilitate comparison.

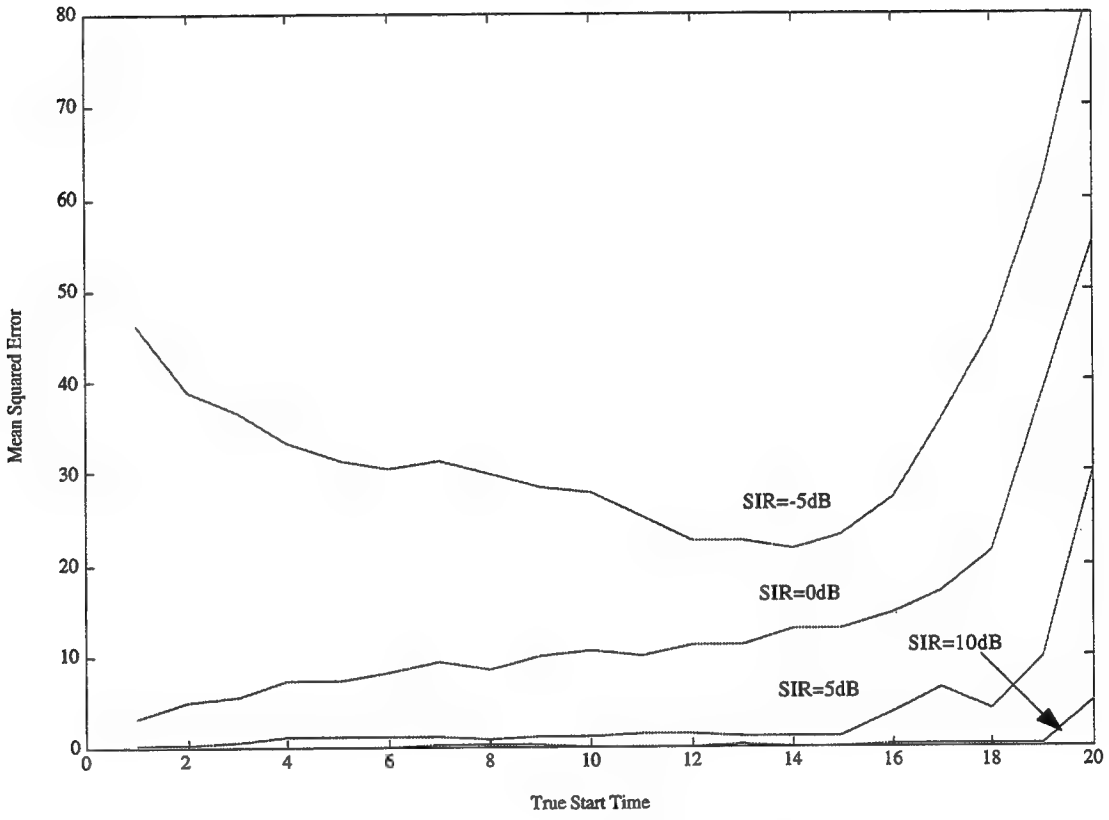

Figure 15: Mean squared error of the onset time estimate for several SIR values as a function of the true start time for $N = 5$ and $M = 20$.

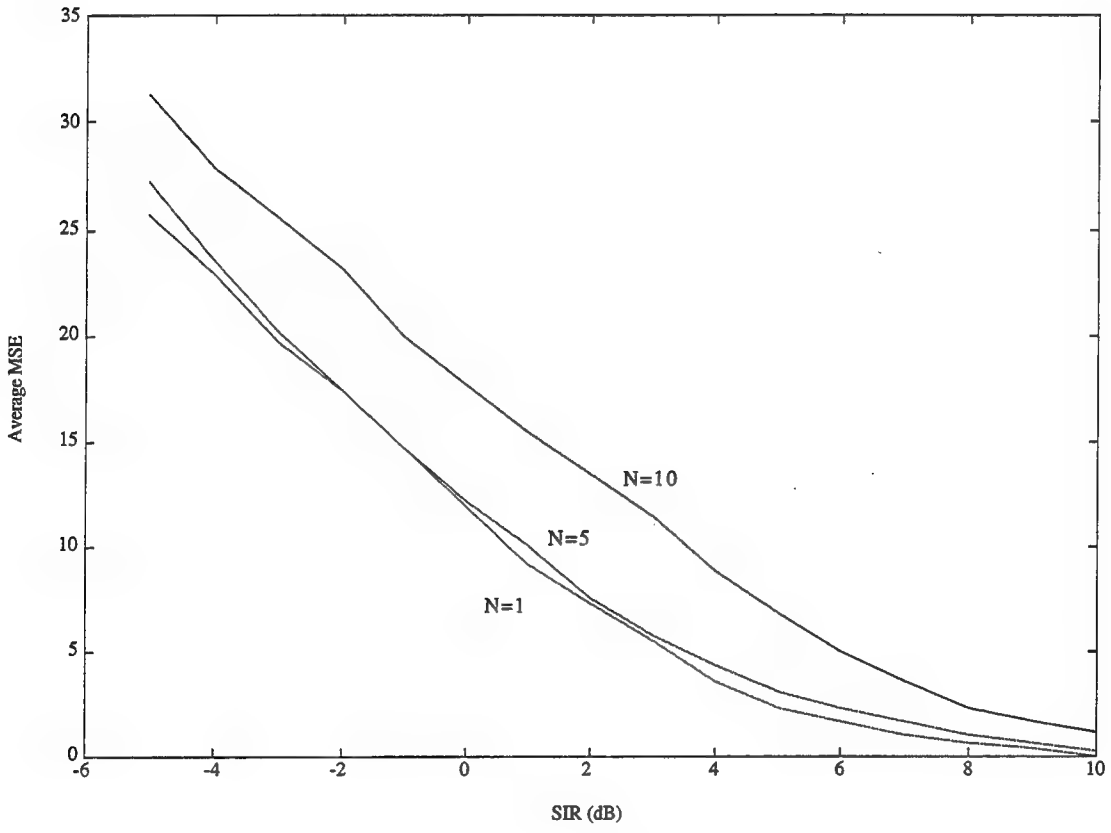

Figure 16: Mean squared error of the onset time estimate averaged over all true start times for $N = 1$, 5, and 10 for $M = 20$ as a function of the signal-to-interference ratio.

Chapter 7

Block–Post Block Simulation

The block detection algorithm proposed and analyzed in chapters 3 - 5 and the post block processing to improve the signal onset time estimate of chapter 6 are applied to simulated array data for a stationary and a non-stationary interference scenario. The array and signal processing consist of a half wavelength, equally spaced, thirty-two sensor line array with frequency domain plane wave beamforming to a broadside look direction. As suggested in chapter 5, beamspace preprocessing is used to reduce the adaptive dimension to $N = 1$ providing conventional beamforming, and to $N = 5$ yielding a potential improvement due to adaptive processing if warranted by the interference scenario. This simulation study will be conducted in two parts: first, the block detection performance will be evaluated by examining histograms of the time between false alarms and the time before detection as well as the average time before detection versus signal strength; second, the combined block detection-post block processing will be

evaluated by histograms of the estimated start time and by considering the mean squared error as a function of the signal strength.

7.1 Interference Scenario

The stationary interference scenario consists of the following: plane wave interferences at 2 degrees and 10 dB, 5 degrees and 10 dB, and -2 degrees and 13 dB, a Butterworth angularly spread interference at 4 degrees and 10 dB with a spread parameter of $\rho = 0.5$, and spatially uncorrelated noise at 0 dB. All angles are from broadside to the array. A scale is applied to this interference scenario so that the conventional beam output interference power is 0 dB at broadside to the array. Thus, evaluating signal strengths ranging from -5 to 5 dB is effectively considering signal-to-interference ratios ranging from -5 to 5 dB for a conventional beamformer ($N = 1$). The non-stationary interference scenario begins with the stationary scenario and has linear changes in the angular position, strength in decibels, and spread of the interference signals as found in table 1. The Butterworth angularly spread interference results in a plane wave when $\rho = 1$ and in spatially uncorrelated noise when $\rho = 0$. The interference scenario is periodic with period equal to 3000 samples. Over one period, the beam output interference power for conventional beamforming ($N = 1$), for beamspace preprocessing to $N = 5$ beams, and for fully adaptive beamforming as a function of time may be found in figure 17. The difference between the conventional beam

output interference power and either adaptive beam output interference power (in dB) is the array gain improvement for the respective adaptive beamformer.

Signal	Time Index	Angle (deg)	Strength (dB)	Spread
1	1	$-2°$	13	1.0
	1000	$-1°$	10	1.0
	2000	$-3°$	15	1.0
	3000	$-2°$	13	1.0
2	1	$2°$	10	1.0
	500	$3°$	5	1.0
	2000	$2°$	13	1.0
	3000	$2°$	10	1.0
3	1	$4°$	10	0.5
	2000	$8°$	15	0.3
	3000	$4°$	10	0.5
4	1	$5°$	10	1.0
	2000	$3°$	10	1.0
	3000	$5°$	10	1.0
5	1	n/a	0	0.0
	3000	n/a	0	0.0

Table 1: Non-stationary interference scenario. The interference parameters change linearly between the specified time indices.

7.2 Block Detection Performance

The GLR locally optimal detector is implemented with an average time between false alarms of $T = 20$ blocks and a block size of $M = 10$ samples. The thresholds were set by using equation 8 of chapter 2. Histograms of the time between false alarms for the stationary and non-stationary interference scenarios are found respectively in figures 18 and 19. The observed average time between false alarms was approximately 25 times the design value for $N = 1$ and 10 times

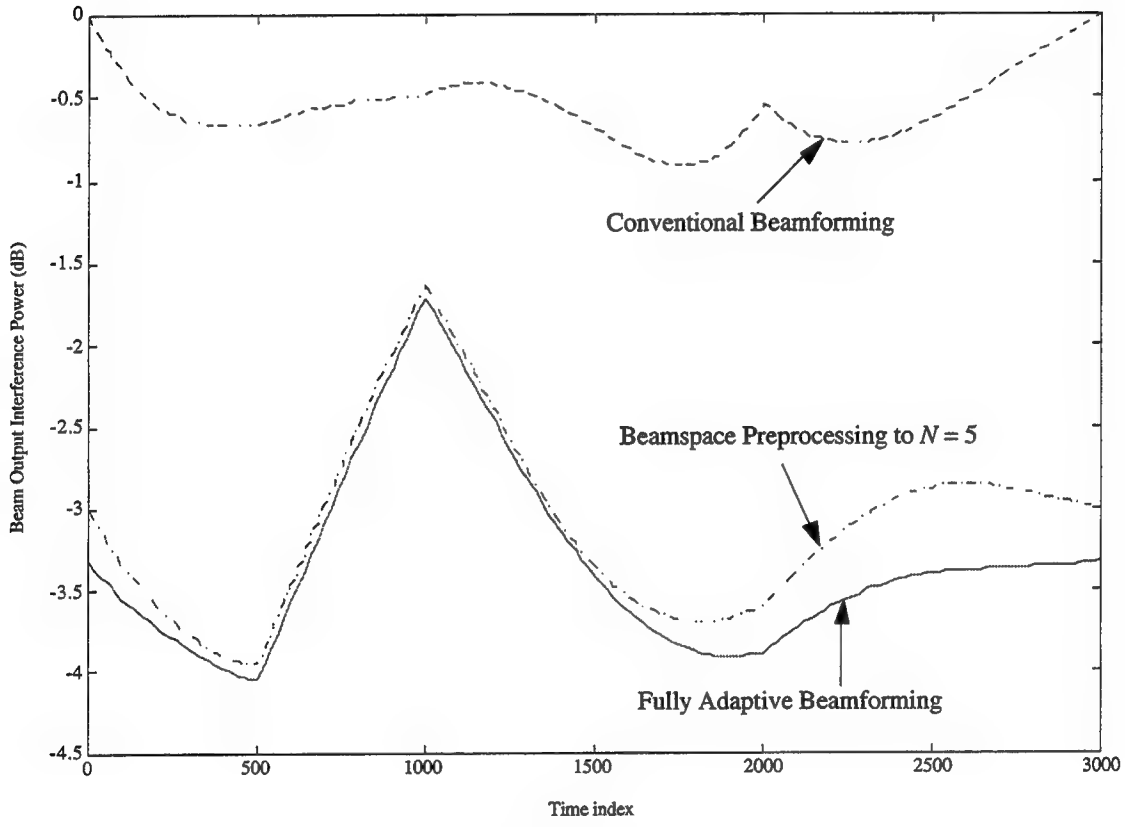

Figure 17: Beam output interference power for conventional beamforming ($N = 1$), for beamspace preprocessing to $N = 5$ beams, and for fully adaptive beamforming for the non-stationary interference scenario.

the design value for $N = 5$. As the design value is only a lower bound, the difference is acceptable, however, tighter bounds would provide more accurate threshold computation. As expected, the time between false alarms was slightly smaller for the non-stationary interference due to the lack of stationarity within each block. The average times between false alarms are still substantially larger than the design value although this may be affected by increasing the block size.

Histograms of the number of blocks before detection for the stationary and non-stationary interference scenarios are found respectively in figures 20 and 21 for the $N = 1$ and $N = 5$ cases for several signal strengths. There seems to be minimal difference between the stationary and non-stationary interference scenarios. As seen in figure 17 the beam output interference power only changes about 0.75 dB during the first 300 time samples (30 blocks) for the $N = 1$ and $N = 5$ case, which may explain the minimal difference. The average number of blocks before detection, as a function of the signal strength, is found in figure 22. Again, the minimal difference between the stationary and non-stationary interference scenarios is observed where the non-stationary case seems to have a slightly smaller average time before detection. This may be accounted for by the increase in the variance of the block level statistics due to the non-stationarity of the interference over one block as in the smaller time between false alarms. The predicted performance, calculated by using the approximation of equation 10 found in chapter 2, is shown in the figure as well, where it is seen that the predicted values are no more than one block below the observed values. The advantage in

preprocessing the array data to a specific dimension is evident in that for signal strengths below 0.5 dB, the $N = 5$ case results in better detection performance while for larger signal strengths, the $N = 1$ case slightly outperforms the $N = 5$ case. This may also be observed in the required SIR curves in figure 11 of chapter 5 where the best preprocessor dimension may change with signal strength.

7.3 Block-Post Block Performance

A block of data thirty samples long immediately previous to the block detection, including the block in which detection occurred, is subjected to the post block processing algorithm of chapter 6. Histograms of the onset time estimate of the combined block-post block processing algorithm for the stationary and non-stationary cases are found respectively in figures 23 and 24 for various signal strengths. The histograms are referenced to the true starting time, which was a randomly selected position within a block. The weaker signal strength cases are clearly biased to the right, which is the result of a delay in the block detection large enough so that the true start time occurred outside of the data segment being processed. Minimal difference is noticed between the stationary and non-stationary interference scenarios.

The mean squared error of the onset time estimate as a function of the signal strength for the stationary and non-stationary interference scenarios and the predicted performance are found in figure 25. The $N = 5$ case produced less mean squared error than the $N = 1$ case as was expected due to the improvement

indicated by the array gain improvement of this interference scenario. The error at low signal strengths is more than an order of magnitude greater than the predicted value, particularly for the $N = 1$ case, as a result of the larger delay in detection previously described. The error for signal strengths greater than 0 dB is on the same order as the predicted values. It is not expected that the error be exact, as the predicted value is the mean squared error averaged over all possible start times which does not exactly represent what is occurring in the simulation. As seen in the histograms of the number of time samples before block detection occurs found in figure 26, the true starting time of the signal is not even approximately uniformly distributed in the segment of data that is applied to the post block processing (the first thirty samples). From this figure it is also clear that at low signal strengths there are a substantial number of trials where the post block processing segment does not include the true start time, leading to an increased mean squared error.

Due to the sequential nature of the block detection algorithm, the average time before detection increases as the signal-to-interference ratio decreases. This has a very debilitating effect on the post block processing leading to substantial errors. Increasing the length of the segment of data submitted to the post block processing may ease this, however, the length is likely to be limited by the stationarity of the interference. An alternative may be choosing the segment of data submitted to the post block processing through the use of an estimate of

the signal-to-interference ratio over the block in which detection occurs and the predicted time before detection.

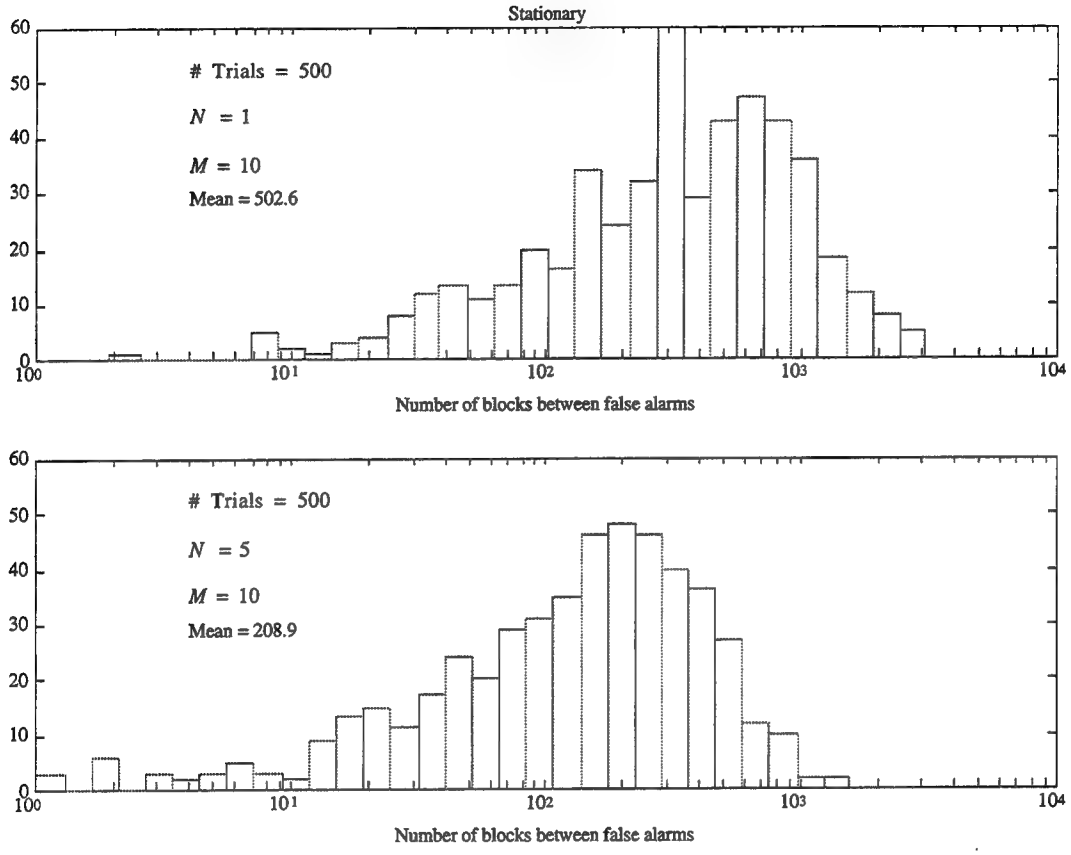

Figure 18: Histograms of the number of blocks between false alarms for $N = 1$ and $N = 5$ with a block size of $M = 10$ for the stationary interference scenario.

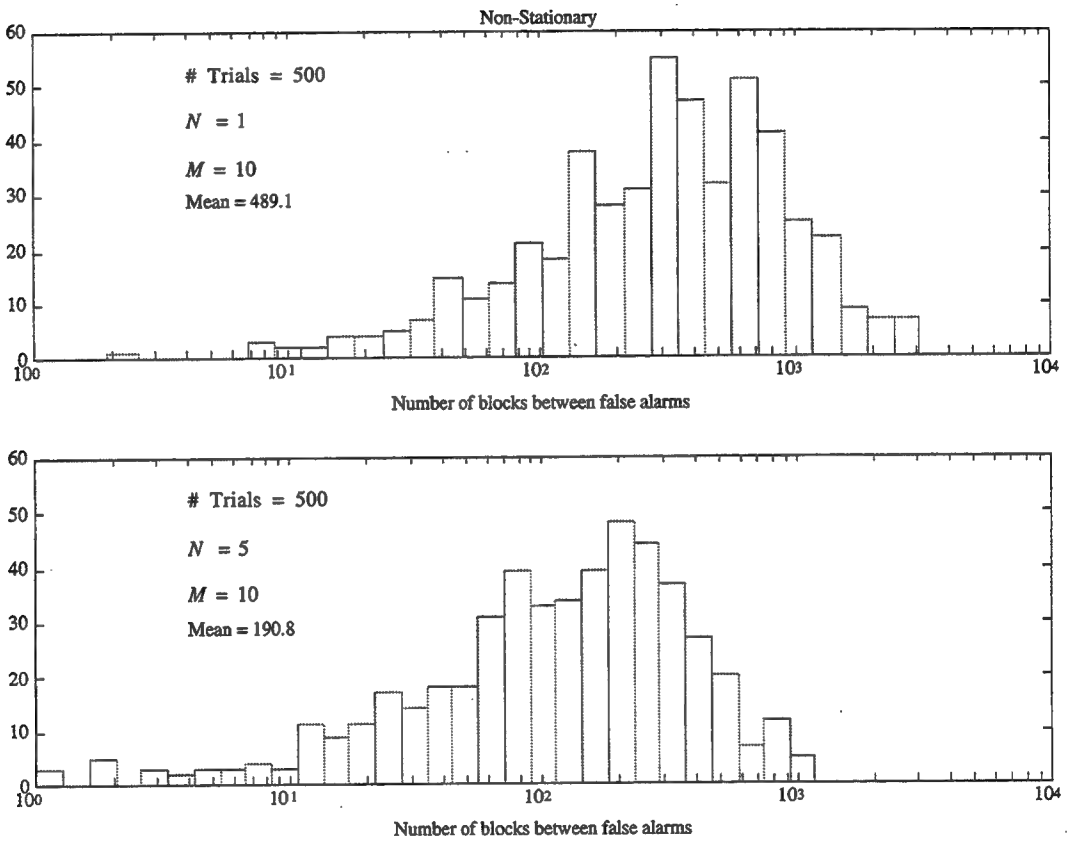

Figure 19: Histograms of the number of blocks between false alarms for $N = 1$ and $N = 5$ with a block size of $M = 10$ for the non-stationary interference scenario.

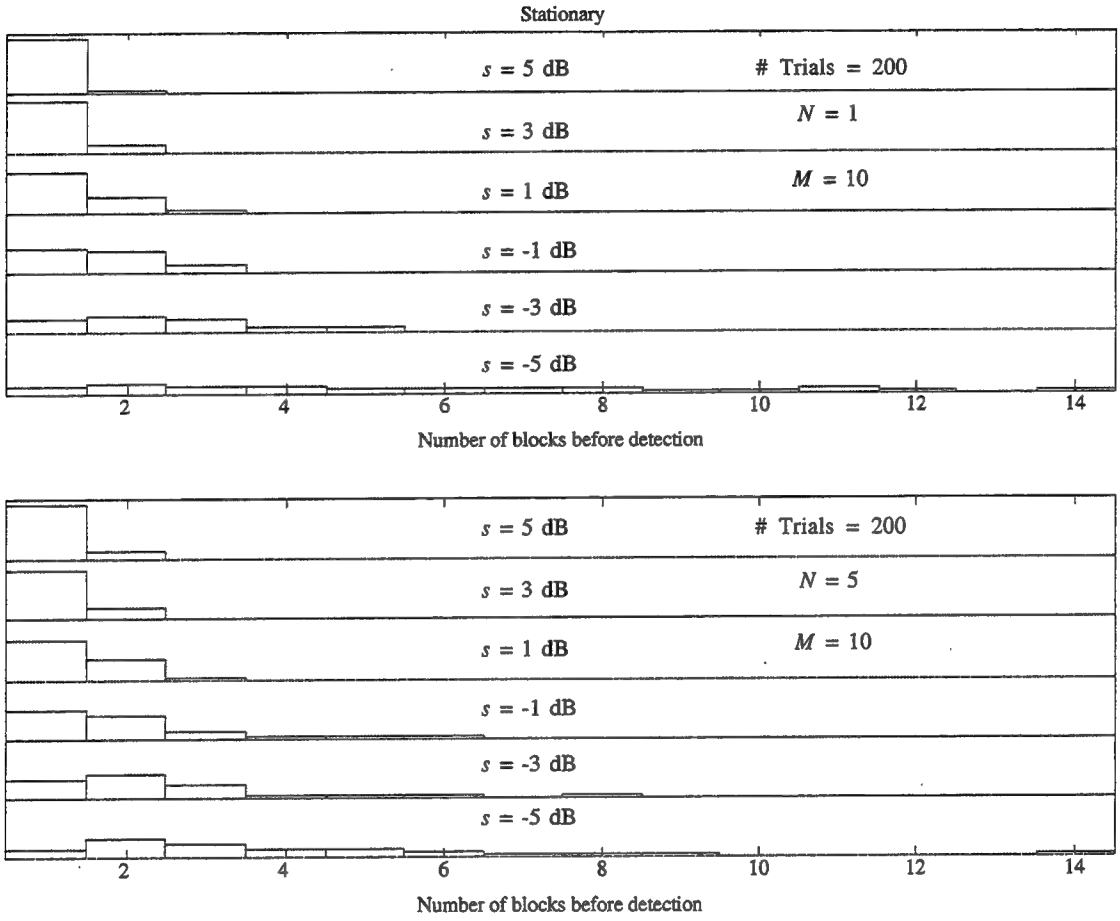

Figure 20: Histograms of the number of blocks before detection for $N = 1$ and $N = 5$ with a block size of $M = 10$ for various signal strengths for the stationary interference scenario. Each histogram has been scaled by the same value to facilitate comparison.

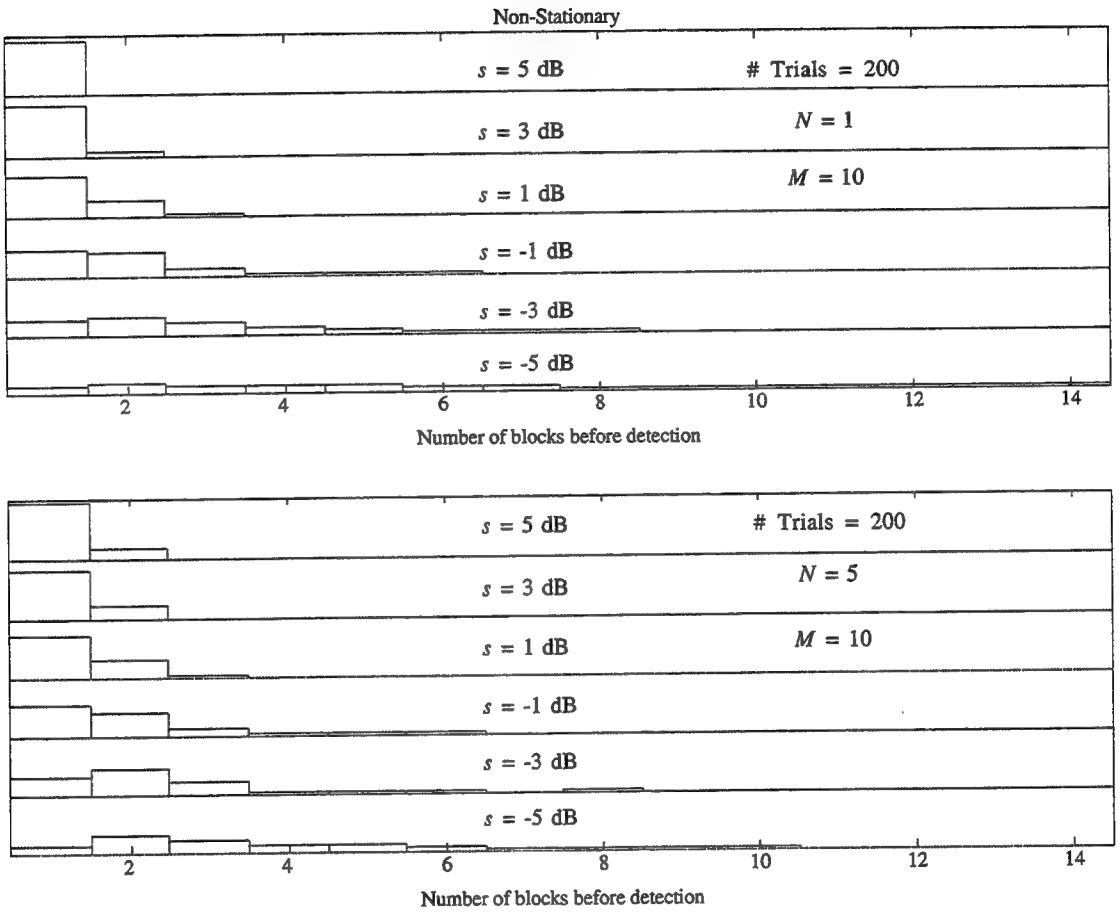

Figure 21: Histograms of the number of blocks before detection for $N = 1$ and $N = 5$ with a block size of $M = 10$ for various signal strengths for the non-stationary interference scenario. Each histogram has been scaled by the same value to facilitate comparison.

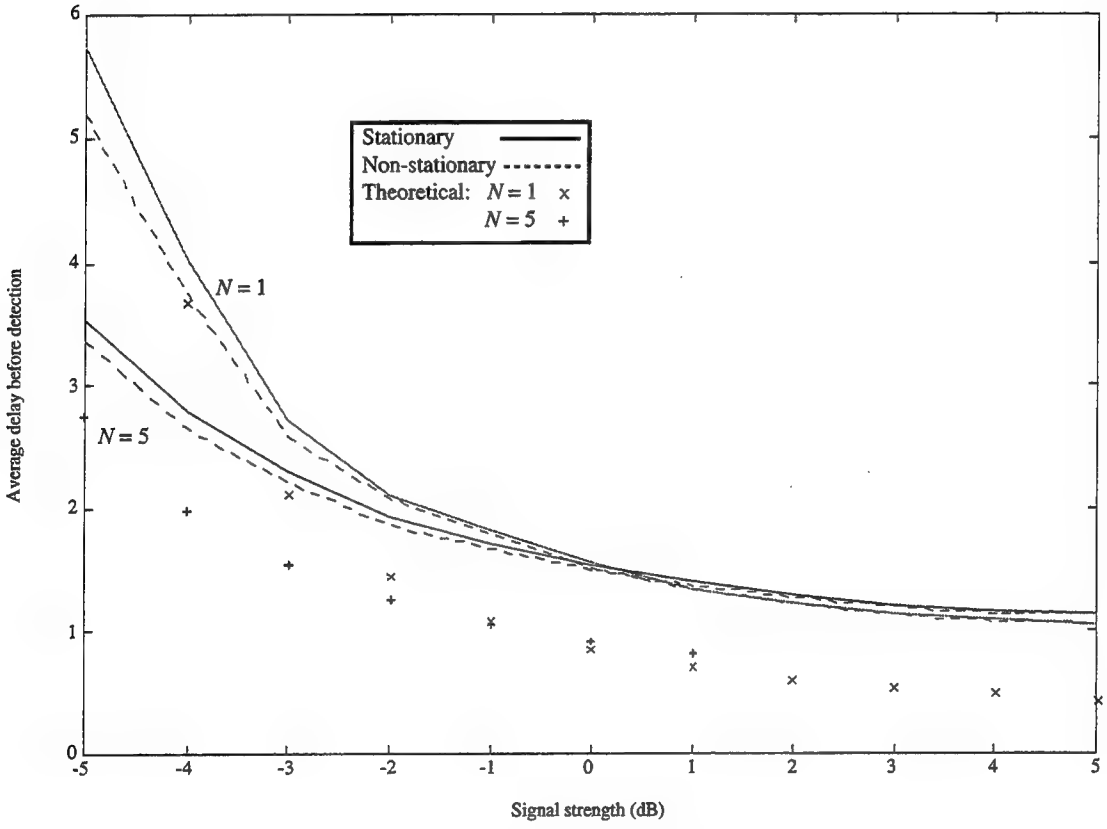

Figure 22: Average number of blocks before detection for $N = 1$ and $N = 5$ with a block size of $M = 10$ as a function of signal strength for the stationary and non-stationary interference scenarios and the predicted performance.

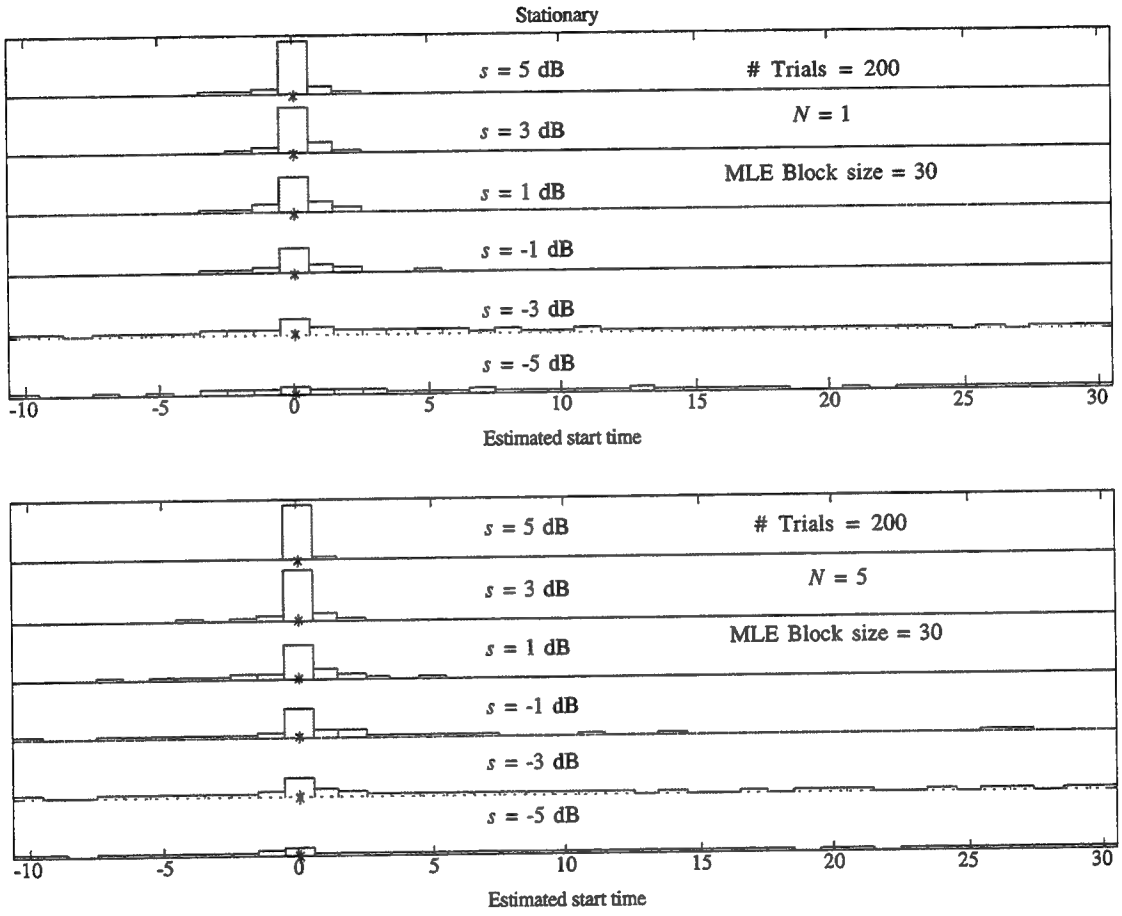

Figure 23: Histograms of the post block, maximum likelihood estimate of the starting time referenced to the true starting time for $N = 1$ and $N = 5$ with a block size of $M = 30$ for various signal strengths for the stationary interference scenario. Each histogram has been scaled by the same value to facilitate comparison.

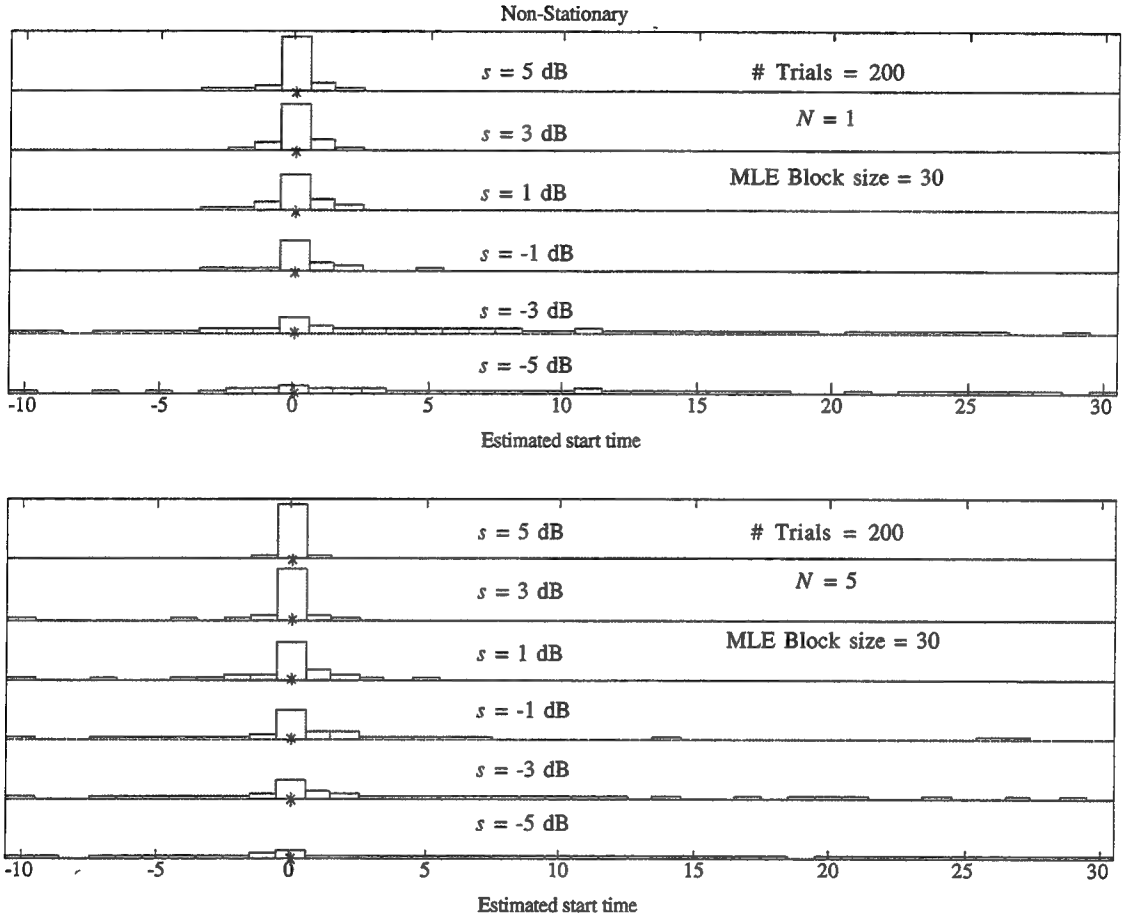

Figure 24: Histograms of the post block, maximum likelihood estimate of the starting time referenced to the true starting time for $N = 1$ and $N = 5$ with a block size of $M = 30$ for various signal strengths for the non-stationary interference scenario. Each histogram has been scaled by the same value to facilitate comparison.

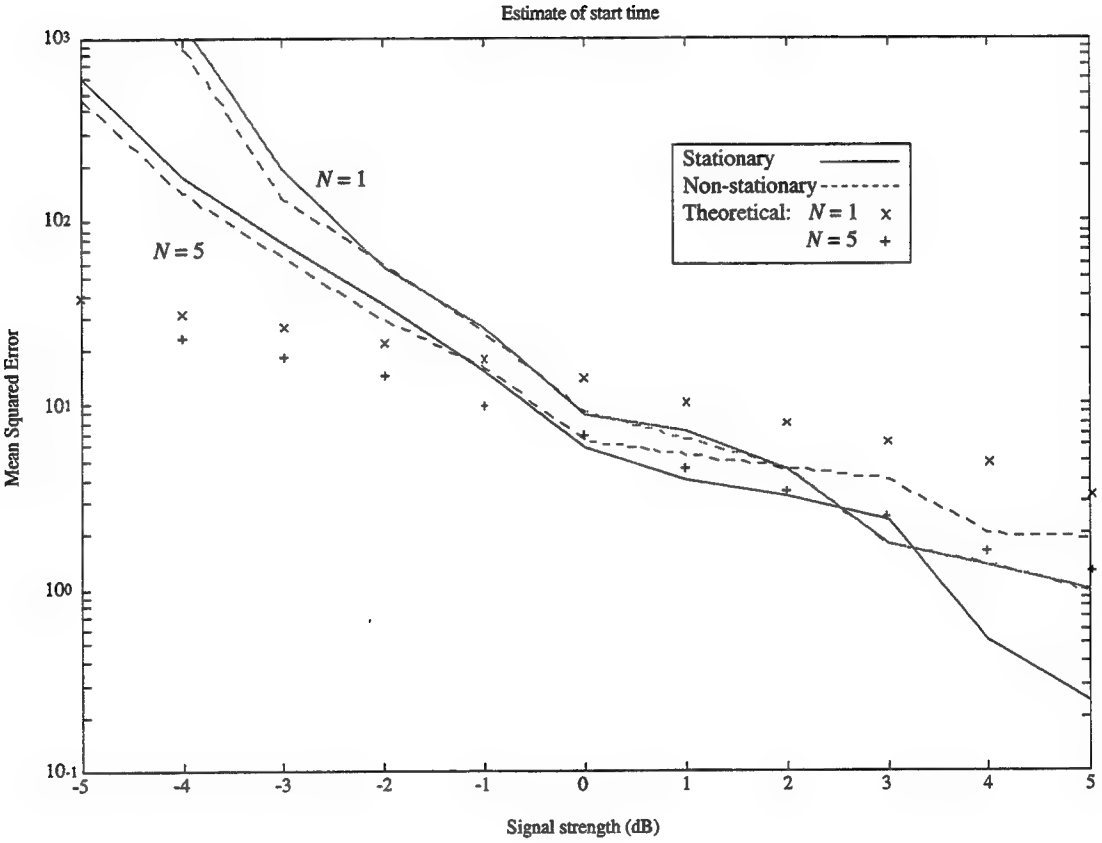

Figure 25: Mean squared error of the post block, maximum likelihood estimate of the starting time for $N = 1$ and $N = 5$ with a block size of $M = 30$ as a function of signal strength for the stationary and non-stationary interference scenarios and the predicted performance.

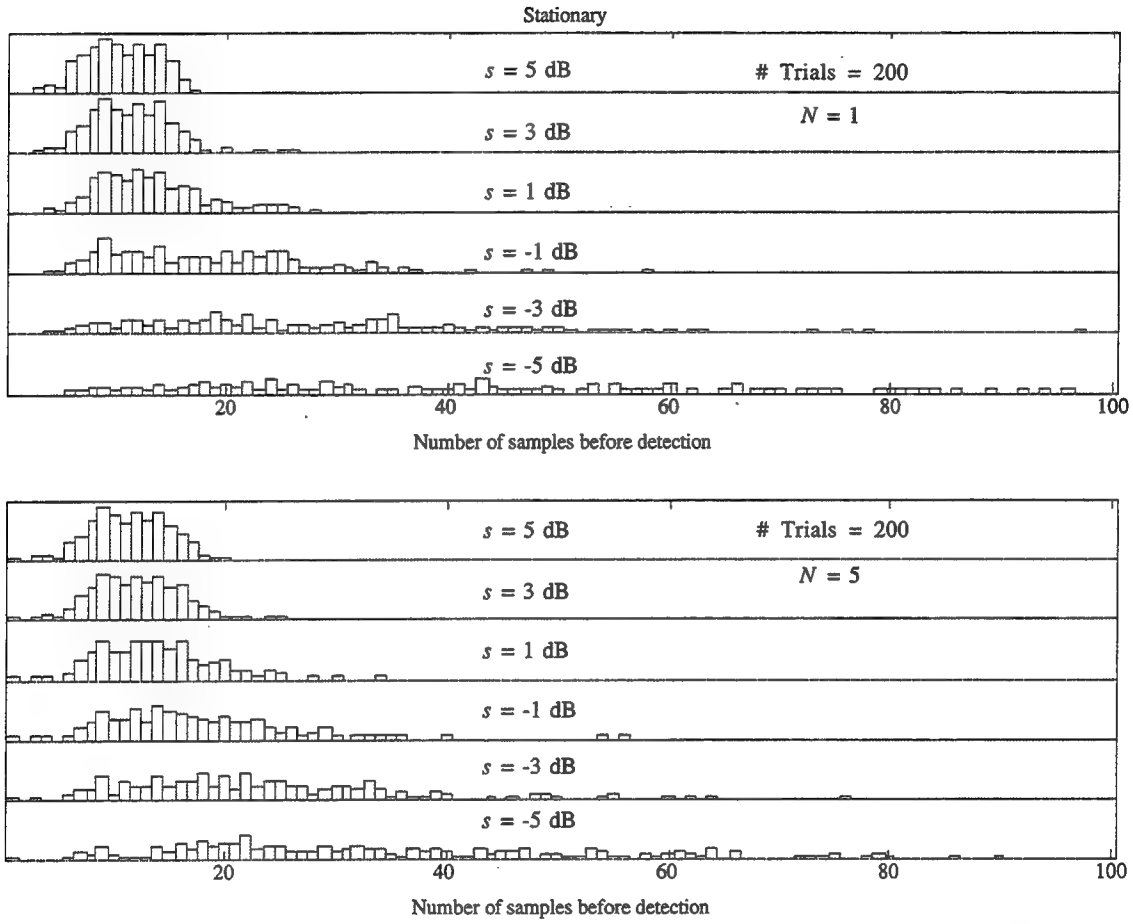

Figure 26: Histogram of the number of time samples before block detection occurs for $N = 1$ and $N = 5$ with a block size of $M = 10$ for various signal strengths for the stationary inteference scenario. Each histogram has been scaled by the same value to facilitate comparison.

Chapter 8

Conclusion

This dissertation has considered the general problem of the determination of the time that a change occurs in the structure or parameterization of the distribution of a sequence of random vectors when some subset of the parameters, nuisance parameters, are common to both the before change (null) hypothesis and the after change (alternative) hypothesis. The proffered solution entails segmenting the data into non-overlapping blocks, from which univariate statistics are formed that are invariant to the nuisance parameters under the null hypothesis and may depend on at most an unknown scalar strength parameter under the alternative hypothesis. The unknown signal strength parameter was treated by applying a locally optimal non-linearity prior to submission to Page's test for the detection of the change. Once Page's test has determined that a change has occurred, post block processing is performed to improve the time resolution of the change time estimation. This structure is appealing for it adequately deals

with non-stationary nuisance parameters, as the block level statistic is invariant to nuisance parameter changes from block to block under the null hypothesis.

A typical radar or sonar application, the detection of the onset of a narrow-band signal of known frequency with unknown amplitude and phase arriving at an array of sensors, was considered. The block detection performance of two block level statistics, a generalized likelihood ratio (GLR) statistic and an adaptive matched filter (AMF) statistic, applied to a log-likelihood ratio for a specified SIR and to a locally optimal non-linearity, was determined as a function of the signal-to-interference ratio. The GLR statistic combined with the locally optimal non-linearity outperformed the other statistic–non-linearity combinations in terms of the asymptotic efficacy. Additionally, the GLR–locally optimal detector has the desirable properties of a closed form detector non-linearity and a standard probability density function, the central chi-squared distribution, when no signal is present. Implementation of the detectors utilizing the locally optimal non-linearity required choosing a bias that insures that the mean of the detector non-linearity is negative when no signal is present. A method for choosing the bias for a general non-linearity that minimizes the signal strength required to achieve a specified asymptotic efficacy, or to equivalently maximize the asymptotic efficacy for a specified signal strength, was derived. The SIR required to achieve a specific asymptotic efficacy was found as a function of the adaptive dimension for the GLR–locally optimal detector. Here it was seen that increasing

the adaptive dimension increases the SIR required to achieve a specific performance level. Preprocessing the array data by a rank reducing, constant, linear transformation provides a method for changing the adaptive dimension of the signal detector algorithm. This also affects the SIR that the detector is subjected to, as a function of the preprocessing structure and the interference scenario. The utility of the required SIR curves was shown by the superposition of the effective SIR achieved by varying the adaptive dimension of beam space preprocessing of the array data for a specific interference scenario. The adaptive dimension providing the SIR that yields the largest asymptotic efficacy was indicated as the best preprocessor for the observed interference scenario.

A post block processing algorithm, in the form of a maximum likelihood estimate of the signal onset time, was derived for the deterministic signal model using the sample level data previous to the block detection, up to the maximum limit of stationarity. The performance of the estimator, in terms of mean squared error, was seen to degrade as the signal-to-interference ratio decreased. Minimal loss was incurred between adaptive dimensions $N = 1$ and $N = 5$, indicating that there may be substantial gain in preprocessing the array data to the dimension appropriate for the specific interference scenario.

The combined sequential block detection-post block processing algorithm was implemented on simulated array data for stationary and non-stationary interference scenarios. The block detection performance indicated that the observed

average time between false alarms was approximately twenty-five times the predicted lower bound for $N = 1$ and ten times the predicted lower bound for $N = 5$ for a block size of $M = 10$. This indicates the need for an improved method of analytically determining the thresholds required to achieve a specified time between false alarms. The non-stationary interference caused a slightly quicker detection along with an associated decrease in the average time between false alarms due to the non-stationarity of the interference within each block. The post block processing algorithm was implemented using the data in the block in which the detection occurred and a finite amount of previous data. For weak signals, the delay in the block detection was often large enough so that the segment of data the post block processing algorithm operated on did not contain the onset of the signal. This indicates that using an estimate of the signal-to-interference ratio to set the block size or the particular segment of data subjected to post block processing may provide improved estimation in the combined algorithm.

This dissertation has explored the use of a block sequential adaptive scheme for the detection of the onset of a narrowband signal at an array of sensors where the signal amplitude and phase and the potentially non-stationary interference covariance matrix are unknown. The block detection algorithm may be easily extended to broadband signals or to narrowband signals with unknown or varying frequency by the appropriate combination of block statistics at several discrete fourier transform frequency bin outputs. The threshold required for implementation may require more, although not substantial, numerical computation. Further

research is also suggested in the following related areas: applying the general algorithm to other array signal models, the method of the block data compression, more accurate bounds for the time between false alarms, and online methods of choosing the data to be submitted to post block processing.

Appendix A

Selected Theorems and Proofs

The following theorems are extensions to the complex case of theorems commonly found in the analysis of multivariate real Gaussian random variables and real Wishart matrices. Background and distributional information on multivariate complex Gaussian random variables and complex Wishart matrices may be found in Goodman [23] and to a lesser extent Anderson [34]. The original theorems may be found in either Searle [25], Muirhead [24], or DeVroye [35]. The extensions are, for the most part, straightforward and result from following the proofs for the real cases. When possible, the proofs have been simplified or deferred.

THEOREM A.1

The following theorem is an extension of Theorem 2, Chapter 2, page 57 of Searle [25] describing the probability density function of certain quadratic forms involving complex Gaussian random vectors.

Theorem A.1 If $\mathbf{x} \sim \mathcal{CN}_n(\mu, \mathbf{\Sigma})$ and $\mathbf{A} = \mathbf{A}^H$, then $2\mathbf{x}^H\mathbf{A}\mathbf{x} \sim \mathcal{X}_r^2(\delta)$ if and only if $\mathbf{A}\mathbf{\Sigma}$ is idempotent. The non-centrality parameter, $\delta = 2\mu^H\mathbf{A}\mu$, and the degrees of freedom parameter, $r = 2\mathrm{tr}(\mathbf{A}\mathbf{\Sigma})$, where $\mathrm{tr}(\mathbf{A}\mathbf{\Sigma})$ is the trace of $\mathbf{A}\mathbf{\Sigma}$.

Proof: For the quadratic form

$$Q = 2\mathbf{x}^H\mathbf{A}\mathbf{x}, \tag{236}$$

consider the moment generating function

$$\begin{aligned} M_Q(t) &= \mathrm{E}\left[e^{tQ}\right] \\ &= \mathrm{E}\left[e^{2t\mathbf{x}^H\mathbf{A}\mathbf{x}}\right] \\ &= \int_{\mathcal{C}^n} \frac{1}{\pi^n |\mathbf{\Sigma}|} \exp\left[2t\mathbf{x}^H\mathbf{A}\mathbf{x} - (\mathbf{x}-\mu)^H \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right] d\mathbf{x} \\ &= \int_{\mathcal{C}^n} \frac{1}{\pi^n |\mathbf{\Sigma}|} \mathrm{e}^V d\mathbf{x}, \end{aligned} \tag{237}$$

where $\mathcal{C}^n$ represents n dimensional complex Euclidean space. The exponent, V, may be massaged into a single quadratic form in $\mathbf{x}$,

$$\begin{aligned} V &= \mathbf{x}^H (2t\mathbf{A})\mathbf{x} - \mathbf{x}^H\boldsymbol{\Sigma}^{-1}\mathbf{x} + \mu^H\boldsymbol{\Sigma}^{-1}\mathbf{x} + \mathbf{x}^H\boldsymbol{\Sigma}^{-1}\mu - \mu^H\boldsymbol{\Sigma}^{-1}\mu \\ &= -(\mathbf{x}-\tilde{\mu})^H\tilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}-\tilde{\mu}) + \tilde{\mu}^H\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mu} - \mu^H\boldsymbol{\Sigma}^{-1}\mu, \end{aligned} \tag{238}$$

where

$$\tilde{\boldsymbol{\Sigma}}^{-1} = \boldsymbol{\Sigma}^{-1} - 2t\mathbf{A} \tag{239}$$

and

$$\tilde{\mu} = \tilde{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\mu. \tag{240}$$

The moment generating function thus becomes,

$$\begin{aligned} M_Q(t) &= \frac{\exp\left[\tilde{\mu}^H\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mu} - \mu^H\boldsymbol{\Sigma}^{-1}\mu\right]}{\left|\tilde{\boldsymbol{\Sigma}}\right|^{-1}|\boldsymbol{\Sigma}|}\int_{\mathcal{C}^n}\frac{\exp\left[-(\mathbf{x}-\tilde{\mu})^H\tilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}-\tilde{\mu})\right]}{\pi^n\left|\tilde{\boldsymbol{\Sigma}}\right|}d\mathbf{x} \\ &= \frac{\exp\left[\mu^H\left(\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\right)\mu\right]}{\left|\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\right|} \\ &= \frac{\exp\left[-\mu^H\left(\mathbf{I}_n - \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}^{-1} - 2t\mathbf{A}\right)^{-1}\right)\boldsymbol{\Sigma}^{-1}\mu\right]}{|\mathbf{I}_n - 2t\mathbf{A}\boldsymbol{\Sigma}|} \\ &= \frac{\exp\left[-\mu^H\left(\mathbf{I}_n - \left(\mathbf{I}_n - 2t\mathbf{A}\boldsymbol{\Sigma}\right)^{-1}\right)\boldsymbol{\Sigma}^{-1}\mu\right]}{|\mathbf{I}_n - 2t\mathbf{A}\boldsymbol{\Sigma}|}, \end{aligned} \tag{241}$$

where $\mathbf{I}_n$ is the n dimensional identity matrix.

Suppose that $\mathbf{A}\boldsymbol{\Sigma}$ is idempotent with rank or trace,

$$\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) = m. \tag{242}$$

Note that the eigenvalues of an idempotent matrix are either zero or one, and, by using a singular value decomposition, the matrix $\mathbf{A}\boldsymbol{\Sigma}$ may be expressed as

$$\mathbf{A}\boldsymbol{\Sigma} = \mathbf{U}\mathbf{V}^H, \tag{243}$$

where the n-by-m matrices $\mathbf{U}$ and $\mathbf{V}$ are orthogonal,

$$\mathbf{V}^H\mathbf{U} = \mathbf{I}_m. \tag{244}$$

Part of the matrix in the quadratic form in the exponent of equation (241) may now be simplified,

$$\begin{aligned} \mathbf{I}_n - \left(\mathbf{I}_n - 2t\mathbf{A}\boldsymbol{\Sigma}\right)^{-1} &= \mathbf{I}_n - \left(\mathbf{I}_n - \mathbf{U}(2t)\mathbf{V}^H\right)^{-1} \\ &= \mathbf{I}_n - \left[\mathbf{I}_n - \mathbf{U}\left(\mathbf{V}^H\mathbf{U} - \frac{1}{2t}\mathbf{I}_m\right)^{-1}\mathbf{V}^H\right] \\ &= \mathbf{U}\left(1 - \frac{1}{2t}\right)^{-1}\mathbf{V}^H \\ &= \left(1 - \frac{1}{1-2t}\right)\mathbf{A}\boldsymbol{\Sigma}. \end{aligned} \tag{245}$$

The determinant in the denominator of equation (241) is clearly

$$|\mathbf{I}_n - 2t\mathbf{A}\boldsymbol{\Sigma}| = (1-2t)^m \tag{246}$$

because $\mathbf{A}\boldsymbol{\Sigma}$ has m non-zero eigenvalues, all equal to one. Substituting (245) and (246) into (241) results in

$$\begin{aligned} M_Q(t) &= \frac{\exp\left[-\left(1-\frac{1}{1-2t}\right)\mu^H\mathbf{A}\mu\right]}{(1-2t)^m} \\ &= \mathrm{e}^{-\mu^H\mathbf{A}\mu}\frac{\exp\left(\frac{1}{1-2t}\mu^H\mathbf{A}\mu\right)}{(1-2t)^{\frac{2m}{2}}}. \end{aligned} \tag{247}$$

The moment generating function of a non-central Chi-squared random variable with r degrees freedom and non-centrality parameter δ, as found in Muirhead [24], is

$$M_{\chi^2_{r,\delta}}(t) = \frac{\mathrm{e}^{-\frac{\delta}{2}}\mathrm{e}^{\frac{\delta}{2(1-2t)}}}{(1-2t)^{\frac{r}{2}}}. \tag{248}$$

Equating equations (247) and (248), it is seen that the quadratic form, Q, has a non-central Chi-squared distribution with $r = 2m = 2\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})$ degrees of freedom and non-centrality parameter $\delta = 2\mu^H\mathbf{A}\mu$.

If it is assumed that the quadratic form has the described non-central Chi-squared distribution, it can be seen that the matrix $\mathbf{A}\boldsymbol{\Sigma}$ must be idempotent with rank equal to $\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) = \frac{r}{2}$ by equating the denominators of equations (241) and (248),

$$\begin{aligned} (1-2t)^{\frac{r}{2}} &= |\mathbf{I}_n - 2t\mathbf{A}\boldsymbol{\Sigma}| \\ &= \prod_{i=1}^{n}(1-2t\lambda_i), \end{aligned} \tag{249}$$

where λ_i are the eigenvalues of $\mathbf{A}\boldsymbol{\Sigma}$. Clearly the left and right sides of equation (249) will be equal only when $\frac{r}{2}$ of the eigenvalues are equal to one with the remaining equal to zero, which results in an idempotent $\mathbf{A}\boldsymbol{\Sigma}$ matrix.

THEOREM A.2

The following theorem is the extension of Theorem 4, Chapter 2, page 59 of Searle [25] showing the independence of quadratic forms involving complex Gaussian random vectors.

Theorem A.2 If $\mathbf{x} \sim \mathcal{CN}_n(\mu, \Sigma)$, $\mathbf{A} = \mathbf{A}^H$ and $\mathbf{B} = \mathbf{B}^H$, then $\mathbf{x}^H\mathbf{A}\mathbf{x}$ and $\mathbf{x}^H\mathbf{B}\mathbf{x}$ are independent if and only if $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{0}$.

Proof: The proof is identical to that of Searle [25] and, as the result is also identical, will not be repeated.

THEOREM A.3

The following theorem is an extension of Theorem 3.2.10 on page 93 of Muirhead [24] describing the probability density functions of the partitions of a complex Wishart matrix and of a particular function of the partitions that enjoys certain independence properties.

Theorem A.3 If the n-by-n matrix $\mathbf{A} \sim \mathcal{CW}_n(m, \mathbf{\Sigma})$ with $\mathbf{A}$ and $\mathbf{\Sigma}$ partitioned into two-by-two blocks,

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \\ \mathbf{\Sigma} &= \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}, \end{aligned} \tag{250}$$

where $\mathbf{A}_{11}$ and $\mathbf{\Sigma}_{11}$ are k-by-k, and if

$$\mathbf{A}_{11\cdot 2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \tag{251}$$

and

$$\mathbf{\Sigma}_{11\cdot 2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}, \tag{252}$$

then

1. $\mathbf{A}_{11\cdot 2} \sim \mathcal{CW}_k(m - n + k, \mathbf{\Sigma}_{11\cdot 2})$ and is independent of $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$.
2. The conditional distribution of $\mathbf{A}_{12}$ given $\mathbf{A}_{22}$ is
$$\mathcal{CN}_{k\times(n-k)}\left(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{A}_{22}, \mathbf{\Sigma}_{11\cdot 2} \otimes \mathbf{A}_{22}\right).$$
3. $\mathbf{A}_{22} \sim \mathcal{CW}_{n-k}(m, \mathbf{\Sigma}_{22})$.

Proof: Consider the probability density function of the complex Wishart distributed matrix, $\mathbf{A}$, as found in Goodman [23],

$$f(\mathbf{A}) = \frac{|\mathbf{A}|^{m-n}}{\Gamma_{m,n}|\mathbf{\Sigma}|^m}\mathrm{etr}\left(-\mathbf{\Sigma}^{-1}\mathbf{A}\right), \tag{253}$$

where 'etr' signifies the matrix exponential trace operation and

$$\Gamma_{m,n} = \pi^{\frac{n(n-1)}{2}}\Gamma(m)\cdots\Gamma(m-n+1) \tag{254}$$

where $\Gamma(x)$ is the standard Gamma function. Applying the k, $n-k$ factorization to the determinants of equation (253) yields

$$\begin{aligned} |\mathbf{A}| &= |\mathbf{A}_{22}| \left|\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right| \\ &= |\mathbf{A}_{22}| \, |\mathbf{A}_{11\cdot 2}| \end{aligned} \tag{255}$$

and, similarly,

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{22}| \, |\boldsymbol{\Sigma}_{11\cdot 2}| \, . \tag{256}$$

The determinant of a partitioned matrix may be found in the Matrix Theory Appendix of Muirhead [24]. The trace of the matrix product $\boldsymbol{\Sigma}^{-1}\mathbf{A}$ in equation (253) must now be massaged into a form containing the matrix partitions $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$ and the matrix $\mathbf{A}_{11\cdot 2}$. This requires substantial algebraic manipulation, beginning with the matrix factorization of the inverse of the matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma}^{-1} = \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} . \tag{257}$$

The following relationships between the partitioned matrix and the partitioned inverse may be easily verified:

$$\mathbf{V}_{11} = \boldsymbol{\Sigma}_{11\cdot 2}^{-1} \tag{258}$$

$$\boldsymbol{\Sigma}_{22}^{-1} = \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} \tag{259}$$

$$\mathbf{V}_{11}^{-1}\mathbf{V}_{12} = -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} . \tag{260}$$

Applying these to the aforementioned trace results in

$$\begin{aligned} \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{A}\right) &= \mathrm{tr}\left(\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\right) \\ &= \mathrm{tr}\left(\begin{bmatrix} \mathbf{V}_{11}\mathbf{A}_{11} + \mathbf{V}_{12}\mathbf{A}_{21} & \mathbf{V}_{11}\mathbf{A}_{12} + \mathbf{V}_{12}\mathbf{A}_{22} \\ \mathbf{V}_{21}\mathbf{A}_{11} + \mathbf{V}_{22}\mathbf{A}_{21} & \mathbf{V}_{21}\mathbf{A}_{12} + \mathbf{V}_{22}\mathbf{A}_{22} \end{bmatrix}\right) \\ &= \mathrm{tr}\left(\mathbf{V}_{11}\mathbf{A}_{11}\right) + \mathrm{tr}\left(\mathbf{V}_{12}\mathbf{A}_{21}\right) + \mathrm{tr}\left(\mathbf{V}_{21}\mathbf{A}_{12}\right) + \mathrm{tr}\left(\mathbf{V}_{22}\mathbf{A}_{22}\right) \\ &= \mathrm{tr}\left(\mathbf{V}_{11}\mathbf{A}_{11}\right) + 2\mathrm{tr}\left(\mathbf{V}_{12}\mathbf{A}_{21}\right) + \mathrm{tr}\left(\mathbf{V}_{22}\mathbf{A}_{22}\right) \\ &= \mathrm{tr}\left(\mathbf{V}_{11}\left[\mathbf{A}_{11\cdot 2} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right]\right) + 2\mathrm{tr}\left(\mathbf{V}_{12}\mathbf{A}_{21}\right) + \mathrm{tr}\left(\mathbf{V}_{22}\mathbf{A}_{22}\right) \\ &= \mathrm{tr}\left(\mathbf{V}_{11}\mathbf{A}_{11\cdot 2}\right) + \mathrm{tr}\left(\mathbf{V}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right) \\ &\quad + 2\mathrm{tr}\left(\mathbf{V}_{12}\mathbf{A}_{21}\right) + \mathrm{tr}\left(\mathbf{V}_{22}\mathbf{A}_{22}\right) \\ &= \mathrm{tr}\left(\boldsymbol{\Sigma}_{11\cdot 2}^{-1}\mathbf{A}_{11\cdot 2}\right) + \mathrm{tr}\left(\mathbf{V}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right) - 2\mathrm{tr}\left(\mathbf{V}_{11}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{21}\right) \\ &\quad + \mathrm{tr}\left(\left[\boldsymbol{\Sigma}_{22}^{-1} + \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\right]\mathbf{A}_{22}\right) \\ &= \mathrm{tr}\left(\boldsymbol{\Sigma}_{11\cdot 2}^{-1}\mathbf{A}_{11\cdot 2}\right) + \mathrm{tr}\left(\mathbf{V}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right) - 2\mathrm{tr}\left(\mathbf{V}_{11}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{21}\right) \\ &\quad + \mathrm{tr}\left(\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right) + \mathrm{tr}\left(\mathbf{A}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathbf{V}_{11}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\right) \end{aligned}$$

$$
\begin{aligned}
&= \operatorname{tr}\left(\boldsymbol{\Sigma}_{11\cdot2}^{-1}\mathbf{A}_{11\cdot2}\right) + \operatorname{tr}\left(\mathbf{V}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right) - \operatorname{tr}\left(\mathbf{V}_{11}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{21}\right) \\
&\quad -\operatorname{tr}\left(\mathbf{V}_{11}\mathbf{A}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right) \\
&\quad +\operatorname{tr}\left(\mathbf{V}_{11}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right) \\
&= \operatorname{tr}\left(\boldsymbol{\Sigma}_{11\cdot2}^{-1}\mathbf{A}_{11\cdot2}\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right) \\
&\quad +\operatorname{tr}\left(\mathbf{V}_{11}\begin{bmatrix} \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \\ -\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \\ +\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\mathbf{A}_{22}^{-1}\mathbf{A}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \end{bmatrix}\right) \\
&= \operatorname{tr}\left(\boldsymbol{\Sigma}_{11\cdot2}^{-1}\mathbf{A}_{11\cdot2}\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right) \\
&\quad +\operatorname{tr}\left(\boldsymbol{\Sigma}_{11\cdot2}^{-1}\left[\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22} - \mathbf{A}_{12}\right]\mathbf{A}_{22}^{-1}\left[\mathbf{A}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \mathbf{A}_{21}\right]\right) \\
&= \operatorname{tr}\left(\boldsymbol{\Sigma}_{11\cdot2}^{-1}\mathbf{A}_{11\cdot2}\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right) \\
&\quad +\operatorname{tr}\left(\boldsymbol{\Sigma}_{11\cdot2}^{-1}\left[\mathbf{A}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right]\mathbf{A}_{22}^{-1}\left[\mathbf{A}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right]^{H}\right)
\end{aligned}
\tag{261}
$$

Substituting equations (255), (256), and (261) into the probability density function of equation (253) yields

$$
\begin{aligned}
f(\mathbf{A}) &= \left\{ \begin{array}{c} \left[\frac{|\mathbf{A}_{11\cdot2}|^{m-n}}{\Gamma_{m,n}|\boldsymbol{\Sigma}_{11\cdot2}|^{m}}\operatorname{etr}\left(-\boldsymbol{\Sigma}_{11\cdot2}^{-1}\mathbf{A}_{11\cdot2}\right)\right]\left[\frac{|\mathbf{A}_{22}|^{m-n}}{|\boldsymbol{\Sigma}_{22}|^{m}}\operatorname{etr}\left(-\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right)\right] \\ \cdot\operatorname{etr}\left(-\boldsymbol{\Sigma}_{11\cdot2}^{-1}\left[\mathbf{A}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right]\mathbf{A}_{22}^{-1}\left[\mathbf{A}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right]^{H}\right) \end{array} \right\} \\
&= \left\{ \left[\frac{|\mathbf{A}_{11\cdot2}|^{m-(n-k)-k}}{\Gamma_{m-(n-k),k}\,|\boldsymbol{\Sigma}_{11\cdot2}|^{m-(n-k)}}\operatorname{etr}\left(-\boldsymbol{\Sigma}_{11\cdot2}^{-1}\mathbf{A}_{11\cdot2}\right)\right]\right. \\
&\quad \cdot\left[\begin{array}{c} \operatorname{etr}\left(-\boldsymbol{\Sigma}_{11\cdot2}^{-1}\left[\mathbf{A}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right]\mathbf{A}_{22}^{-1}\left[\mathbf{A}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right]^{H}\right) \\ \frac{\Gamma_{m-(n-k),k}\Gamma_{m,n-k}}{\Gamma_{m,n}}\frac{|\mathbf{A}_{22}|^{-k}}{|\boldsymbol{\Sigma}_{11\cdot2}|^{n-k}} \end{array}\right] \\
&\quad \left.\cdot\left[\frac{|\mathbf{A}_{22}|^{m-(n-k)}}{\Gamma_{m,n-k}\,|\boldsymbol{\Sigma}_{22}|^{m}}\operatorname{etr}\left(-\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right)\right]\right\} \\
&= f(\mathbf{A}_{11\cdot2})\,f(\mathbf{A}_{12}|\mathbf{A}_{22})\,f(\mathbf{A}_{22}),
\end{aligned}
\tag{262}
$$

where it is recognized that $\mathbf{A}_{11\cdot2}$ and $\mathbf{A}_{22}$ have complex Wishart distributions,

$$
\mathbf{A}_{11\cdot2} \sim \mathcal{CW}_k\left(m-(n-k), \boldsymbol{\Sigma}_{11\cdot2}\right) \tag{263}
$$

and

$$
\mathbf{A}_{22} \sim \mathcal{CW}_{n-k}\left(m, \boldsymbol{\Sigma}_{22}\right), \tag{264}
$$

thus proving the first and third parts of the theorem. The independence of $\mathbf{A}_{11\cdot2}$ from $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$ is seen from the factorization of the probability density function

of equation (262). The probability density function of $\mathbf{A}_{12}$ conditioned on $\mathbf{A}_{22}$ may be simplified by considering

$$\begin{aligned}\frac{\Gamma_{m-(n-k),k}\Gamma_{m,n-k}}{\Gamma_{m,n}} &= \left\{ \begin{array}{c} \frac{\Gamma(m-n+k)\cdots\Gamma(m-n+1)\cdot\Gamma(m)\cdots\Gamma(m-n+k+1)}{\Gamma(m)\cdots\Gamma(m-n+1)} \\ \cdot\pi^{\frac{1}{2}[k(k-1)+(n-k)(n-k-1)-n(n-1)]} \end{array} \right\} \\ &= \pi^{\frac{1}{2}\left[k^2-k+n^2-2nk+k^2-n+k-n^2+n\right]} \\ &= \pi^{-k(n-k)}. \end{aligned} \tag{265}$$

Substituting equation (265) into the conditional probability density function of $\mathbf{A}_{12}$ yields

$$f(\mathbf{A}_{12}|\mathbf{A}_{22}) = \frac{\operatorname{etr}\left(-\boldsymbol{\Sigma}_{11\cdot 2}^{-1}\left[\mathbf{A}_{12}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right]\mathbf{A}_{22}^{-1}\left[\mathbf{A}_{12}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}\right]^H\right)}{\pi^{k(n-k)}\left|\mathbf{A}_{22}\right|^k\left|\boldsymbol{\Sigma}_{11\cdot 2}\right|^{n-k}} \tag{266}$$

which is recognized as the probability density function of a complex Gaussian random matrix,

$$\mathbf{A}_{12}|\mathbf{A}_{22} \sim \mathcal{CN}_{k\times(n-k)}\left(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{A}_{22}, \boldsymbol{\Sigma}_{11\cdot 2}\otimes\mathbf{A}_{22}\right), \tag{267}$$

proving the second part of the theorem.

THEOREM A.4

The following theorem is an extension of Theorem 3.2.11 on page 95 of Muirhead [24] describing the probability density function of the inverse of a matrix quadratic form involving the inverse of a complex Wishart distributed matrix.

Theorem A.4 If $\mathbf{A} \sim \mathcal{CW}_n(m, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ is full rank, and $\mathbf{P}$ is a k-by-n matrix with rank k, then

$$\left(\mathbf{P}\mathbf{A}^{-1}\mathbf{P}^H\right)^{-1} \sim \mathcal{CW}_k\left(m-n+k, \left(\mathbf{P}\boldsymbol{\Sigma}^{-1}\mathbf{P}^H\right)^{-1}\right). \tag{268}$$

Proof: If $\boldsymbol{\Sigma}$ is full rank, it may be factored into

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^H, \tag{269}$$

where $\boldsymbol{\Gamma}$ is also of full rank. Set

$$\mathbf{B} = \boldsymbol{\Gamma}^{-1}\mathbf{A}\left(\boldsymbol{\Gamma}^H\right)^{-1}. \tag{270}$$

Since $\boldsymbol{\Gamma}$ is constant and of full rank, $\mathbf{B}$ is distributed as

$$\mathbf{B} \sim \mathcal{CW}_n(m, \mathbf{I}_n). \tag{271}$$

Define the k-by-n, rank k matrix

$$\mathbf{R} = \mathbf{P}\left(\mathbf{\Gamma}^H\right)^{-1}. \tag{272}$$

Substituting equations (272) and (270) into the matrix product described in the theorem results in

$$\begin{aligned}\left(\mathbf{P}\mathbf{A}^{-1}\mathbf{P}^H\right)^{-1} &= \left(\mathbf{R}\mathbf{\Gamma}^H\mathbf{A}^{-1}\mathbf{\Gamma}\mathbf{R}^H\right)^{-1} \\ &= \left(\mathbf{R}\left[\mathbf{\Gamma}^{-1}\mathbf{A}\left(\mathbf{\Gamma}^H\right)^{-1}\right]^{-1}\mathbf{R}^H\right)^{-1} \\ &= \left(\mathbf{R}\mathbf{B}^{-1}\mathbf{R}^H\right)^{-1}. \end{aligned} \tag{273}$$

Similarly,

$$\begin{aligned}\left(\mathbf{P}\mathbf{\Sigma}^{-1}\mathbf{P}^H\right)^{-1} &= \left(\mathbf{R}\mathbf{\Gamma}^H\mathbf{\Sigma}^{-1}\mathbf{\Gamma}\mathbf{R}^H\right)^{-1} \\ &= \left[\mathbf{R}\left(\mathbf{\Gamma}^{-1}\mathbf{\Sigma}\left(\mathbf{\Gamma}^H\right)^{-1}\right)^{-1}\mathbf{R}^H\right]^{-1} \\ &= \left(\mathbf{R}\mathbf{R}^H\right)^{-1}. \end{aligned} \tag{274}$$

Using equations (273) and (274), it is seen that the theorem, as stated in equation (268), simplifies to showing that

$$\left(\mathbf{R}\mathbf{B}^{-1}\mathbf{R}^H\right)^{-1} \sim \mathcal{CW}_n\left(m-n+k, \left(\mathbf{R}\mathbf{R}^H\right)^{-1}\right). \tag{275}$$

Using a singular value decomposition, the matrix $\mathbf{R}$ may be factored into

$$\mathbf{R} = \mathbf{U}\left[\begin{array}{cc}\mathbf{I}_k & \mathbf{0}\end{array}\right]\mathbf{V}^H, \tag{276}$$

where $\mathbf{U}$ is k-by-k and non-singular and $\mathbf{V}$ is n-by-n and orthogonal,

$$\mathbf{V}^{-1} = \mathbf{V}^H. \tag{277}$$

Substituting this factorization into equations (273) and (274) results in

$$\begin{aligned}\left(\mathbf{R}\mathbf{B}^{-1}\mathbf{R}^H\right)^{-1} &= \left(\mathbf{U}\left[\begin{array}{cc}\mathbf{I}_k & \mathbf{0}\end{array}\right]\mathbf{V}^H\mathbf{B}^{-1}\mathbf{V}\left[\begin{array}{c}\mathbf{I}_k \\ \mathbf{0}\end{array}\right]\mathbf{U}^H\right)^{-1} \\ &= \left(\mathbf{U}^H\right)^{-1}\left(\left[\begin{array}{cc}\mathbf{I}_k & \mathbf{0}\end{array}\right]\left(\mathbf{V}^H\mathbf{B}\mathbf{V}\right)^{-1}\left[\begin{array}{c}\mathbf{I}_k \\ \mathbf{0}\end{array}\right]\right)^{-1}\mathbf{U}^{-1} \\ &= \left(\mathbf{U}^H\right)^{-1}\left(\left[\begin{array}{cc}\mathbf{I}_k & \mathbf{0}\end{array}\right]\mathbf{C}^{-1}\left[\begin{array}{c}\mathbf{I}_k \\ \mathbf{0}\end{array}\right]\right)^{-1}\mathbf{U}^{-1}, \end{aligned} \tag{278}$$

where

$$\mathbf{C} = \mathbf{V}^H\mathbf{B}\mathbf{V}, \tag{279}$$

and

$$\begin{aligned}\left(\mathbf{R}\mathbf{R}^H\right)^{-1} &= \left(\mathbf{U}\left[\begin{array}{cc}\mathbf{I}_k & \mathbf{0}\end{array}\right]\mathbf{V}^H\mathbf{V}\left[\begin{array}{c}\mathbf{I}_k \\ \mathbf{0}\end{array}\right]\mathbf{U}^H\right)^{-1} \\ &= \left(\mathbf{U}\left[\begin{array}{cc}\mathbf{I}_k & \mathbf{0}\end{array}\right]\left[\begin{array}{c}\mathbf{I}_k \\ \mathbf{0}\end{array}\right]\mathbf{U}^H\right)^{-1} \\ &= \left(\mathbf{U}\mathbf{U}^H\right)^{-1}. \end{aligned} \tag{280}$$

Let

$$\mathbf{C} = \left[\begin{array}{cc}\mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22}\end{array}\right] \tag{281}$$

and

$$\mathbf{C}^{-1} = \mathbf{D} = \left[\begin{array}{cc}\mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_{22}\end{array}\right] \tag{282}$$

be k-by-$n-k$ partitions of $\mathbf{C}$ and $\mathbf{D} = \mathbf{C}^{-1}$. Then, equation (278) becomes

$$\begin{aligned}\left(\mathbf{R}\mathbf{B}^{-1}\mathbf{R}^H\right)^{-1} &= \left(\mathbf{U}^H\right)^{-1}\mathbf{D}_{11}^{-1}\mathbf{U}^{-1} \\ &= \left(\mathbf{U}^H\right)^{-1}\mathbf{C}_{11\cdot 2}\mathbf{U}^{-1}, \end{aligned} \tag{283}$$

where $\mathbf{C}_{11\cdot 2}$ is as defined in Theorem A.3 and, as seen in Muirhead [24], is equal to $\mathbf{D}_{11}^{-1}$. Since $\mathbf{V}$ is orthogonal, $\mathbf{C}$ is distributed as

$$\mathbf{C} \sim \mathcal{CW}_n\left(m, \mathbf{I}_n\right). \tag{284}$$

Applying part (i) of Theorem A.3, it is seen that

$$\mathbf{C}_{11\cdot 2} \sim \mathcal{CW}_k\left(m-n+k, \mathbf{I}_k\right), \tag{285}$$

which, when applied to equation (283), results in

$$\begin{aligned}\left(\mathbf{R}\mathbf{B}^{-1}\mathbf{R}^H\right)^{-1} &\sim \mathcal{CW}_k\left(m-n+k, \left(\mathbf{U}\mathbf{U}^H\right)^{-1}\right) \\ &\sim \mathcal{CW}_k\left(m-n+k, \left(\mathbf{R}\mathbf{R}^H\right)^{-1}\right), \end{aligned} \tag{286}$$

which completes the proof.

THEOREM A.5

The following theorem is an extension of Theorem 3.2.12 on page 96 of Muirhead [24] describing the probability density function of the ratio of quadratic forms involving the inverse of the scale matrix and the inverse of a random sample of a complex Wishart distribution. It is interesting to note that the resulting distribution does not depend on the vector in the quadratic forms if it is independent of the complex Wishart distributed matrix.

Theorem A.5 If $\mathbf{A} \sim \mathcal{CW}_n(m, \mathbf{\Sigma})$ where $m > n-1$ and if $\mathbf{y}$ is any n-by-1 random vector independent of $\mathbf{A}$ such that $Pr(\mathbf{y} = \mathbf{0}) = 0$, then

$$2\frac{\mathbf{y}^H\mathbf{\Sigma}^{-1}\mathbf{y}}{\mathbf{y}^H\mathbf{A}^{-1}\mathbf{y}} \sim \mathcal{X}^2_{2(m-n+1)}, \tag{287}$$

and is independent of $\mathbf{y}$.

Proof: In Theorem A.4, let $\mathbf{P} = \mathbf{y}^H$. Then,

$$W = \left(\mathbf{y}^H\mathbf{A}^{-1}\mathbf{y}\right)^{-1} \sim \mathcal{CW}_1\left(m-n+1, \left(\mathbf{y}^H\mathbf{\Sigma}^{-1}\mathbf{y}\right)^{-1}\right), \tag{288}$$

which has probability density function,

$$\begin{aligned} f_W(w) &= \frac{|w|^{m-n+1-1}\,\mathrm{etr}\left(-\frac{w}{\theta}\right)}{\Gamma(m-n+1)\,|\theta|^{m-n+1}} \\ &= \frac{w^{m-n}\mathrm{e}^{-\frac{w}{\theta}}}{\Gamma(m-n+1)\,\theta^{m-n+1}}, \end{aligned} \tag{289}$$

where

$$\theta = \left(\mathbf{y}^H\mathbf{\Sigma}^{-1}\mathbf{y}\right)^{-1}. \tag{290}$$

Performing the transformation

$$Z = \frac{2}{\theta}W = 2\frac{\mathbf{y}^H\mathbf{\Sigma}^{-1}\mathbf{y}}{\mathbf{y}^H\mathbf{A}^{-1}\mathbf{y}}, \tag{291}$$

yields the desired scaled ratio of quadratic forms in equation (287). The probability density function of Z is found to be

$$\begin{aligned} f_Z(z) &= \frac{\theta\left(\frac{z\theta}{2}\right)^{m-n}\mathrm{e}^{-\frac{z}{2}}}{\Gamma(m-n+1)\,2\theta^{m-n+1}} \\ &= \frac{z^{m-n+1-1}\mathrm{e}^{-\frac{z}{2}}}{\Gamma(m-n+1)\,2^{m-n+1}} \\ &= \frac{z^{\frac{r}{2}-1}\mathrm{e}^{-\frac{z}{2}}}{\Gamma\left(\frac{r}{2}\right)2^{\frac{r}{2}}}, \end{aligned} \tag{292}$$

where

$$r = 2(m-n+1), \tag{293}$$

which is a central Chi-squared distribution with $2(m-n+1)$ degrees of freedom.

The following theorem is based on a result found in DeVroye [35].

Theorem A.6 If $W \sim \mathcal{X}_n^2$ and $V \sim \mathcal{X}_m^2$ are independent then

$$Z = \frac{W}{W+V} \sim \beta eta\left(\frac{n}{2}, \frac{m}{2}\right).$$

Proof: Consider the transformation

$$(W, V) \rightarrow (Z, Y), \tag{294}$$

where

$$Y = V \tag{295}$$

and

$$Z = \frac{W}{W+V}. \tag{296}$$

The range of Y and Z are

$$0 \leq Y \leq \infty \tag{297}$$

and

$$0 \leq Z \leq 1. \tag{298}$$

Inverting the above one-to-one transformation yields

$$V = Y \tag{299}$$

and

$$W = \frac{-ZY}{Z-1}. \tag{300}$$

The Jacobian matrix of the transformation, the partial derivatives of (W, V) with respect to (Z, Y), is

$$\mathbf{J} = \begin{bmatrix} \frac{Y}{(Z-1)^2} & \frac{-Z}{Z-1} \\ 0 & 1 \end{bmatrix}. \tag{301}$$

The absolute value of the determinant of the Jacobian matrix is

$$||\mathbf{J}|| = \frac{Y}{(Z-1)^2}. \tag{302}$$

The joint probability density function of Z and Y has the form

$$f_{Z,Y}(z,y) = f_{V,W}\left(y, -\frac{zy}{z-1}\right)\|\mathbf{J}\|. \tag{303}$$

If $W \sim \mathcal{X}_n^2$ and $V \sim \mathcal{X}_m^2$, and are independent, their joint probability density function is

$$f_{V,W}(v,w) = \left(\frac{v^{\frac{m}{2}-1}e^{-\frac{v}{2}}}{\Gamma(m)\,2^{\frac{m}{2}}}\right)\left(\frac{w^{\frac{n}{2}-1}e^{-\frac{w}{2}}}{\Gamma(n)\,2^{\frac{n}{2}}}\right). \tag{304}$$

Substituting equation (304) into (303) results in

$$\begin{aligned} f_{Z,Y}(z,y) &= \left(\frac{y^{\frac{m}{2}-1}e^{-\frac{y}{2}}}{\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2}}}\right)\left(\frac{\left(-\frac{zy}{z-1}\right)^{\frac{n}{2}-1}e^{\frac{zy}{2(z-1)}}}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}}\right)\left(\frac{y}{(z-1)^2}\right) \\ &= \frac{z^{\frac{n}{2}-1}(1-z)^{-\frac{n}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{\frac{n+m}{2}}}\left(y^{\frac{m+n}{2}-1}\mathrm{e}^{\frac{-y}{2(1-z)}}\right) \\ &= \frac{z^{\frac{n}{2}-1}(1-z)^{-\frac{n}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{\frac{n+m}{2}}}\left(y^{p-1}\mathrm{e}^{-\frac{y}{\theta}}\right), \end{aligned} \tag{305}$$

where

$$p = \frac{m+n}{2} \tag{306}$$

and

$$\theta = 2(1-z). \tag{307}$$

The probability density function for Z is found by marginalizing over Y,

$$\begin{aligned} f_Z(z) &= \int_{y=0}^{\infty} f_{Z,Y}(z,y)\,dy \\ &= \frac{z^{\frac{n}{2}-1}(1-z)^{-\frac{n}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{\frac{n+m}{2}}}\Gamma(p)\,\theta^p\int_{y=0}^{\infty}\frac{y^{p-1}\mathrm{e}^{-\frac{y}{\theta}}}{\Gamma(p)\,\theta^p}dy \\ &= \frac{z^{\frac{n}{2}-1}(1-z)^{-\frac{n}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{\frac{n+m}{2}}}\Gamma\left(\frac{n+m}{2}\right)(2(1-z))^{\frac{n+m}{2}} \\ &= \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}z^{\frac{n}{2}-1}(1-z)^{\frac{m}{2}-1}, \end{aligned} \tag{308}$$

where it is recognized that

$$\int_{y=0}^{\infty}\frac{y^{p-1}\mathrm{e}^{-\frac{y}{\theta}}}{\Gamma(p)\,\theta^p}dy = 1 \tag{309}$$

because it is the integral of the probability density function of a standard Gamma random variable with p degrees of freedom and scale θ. In its final form, equation (308) is recognized as a standard Beta probability density function with parameters $\frac{n}{2}$ and $\frac{m}{2}$.

Appendix B

Non-Central F Distribution Approximation

As seen in appendix C, the non-central Fisher's f density function requires the evaluation of an infinite summation. As evidenced by Tiku in [26], there exist several methods for approximating the upper tail probability of this distribution. The *Three-Moment Approximation* is indicated as the most reasonable in terms of computational requirements and accuracy. The probability density function may be evaluated by differentiating and negating the upper tail probability. The following outlines this procedure as applied to Tiku's Three-Moment Approximation utilizing Streit's [36] description of the upper tail probability approximation.

B.1 Upper Tail Probability

Let the upper tail probability of a non-central Fisher's f distribution with n_1 and n_2 degrees of freedom and non-centrality parameter δ be defined by

$$Q\left(x|n_1,n_2,\delta\right)=\int_{z=x}^{\infty}f\left(z|n_1,n_2,\delta\right)dz. \tag{310}$$

Tiku's Three Moment Approximation states that

$$Q\left(x|n_1,n_2,\delta\right)\approx I\left(y|\frac{n_2}{2},\frac{b}{2}\right), \tag{311}$$

where $I\left(y|\alpha,\beta\right)$ describes the incomplete Gamma function

$$I\left(x|\alpha,\beta\right)=\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}\int_{t=0}^{x}t^{\alpha-1}\left(1-t\right)^{\beta-1}dt, \tag{312}$$

and

$$y \;=\; \left[1+\frac{b}{n_2h}\left(x+c\right)\right]^{-1} \tag{313}$$

$$b = \frac{n_2-2}{2}\left[\sqrt{\frac{E}{E-4}}-1\right] \tag{314}$$

$$h = \frac{bH}{n_1K\left(2b+n_2-2\right)} \tag{315}$$

$$c = \frac{n_2}{n_2-2}\left(h-1-\frac{\delta}{n_1}\right) \tag{316}$$

$$H = \left\{\begin{array}{c} 2\left(n_1+\delta\right)^3+\left(n_1+3\delta\right)\left(n_2-2\right)^2 \\ +3\left(n_1+\delta\right)\left(n_1+2\delta\right)\left(n_2-2\right) \end{array}\right\} \tag{317}$$

$$K = \left(n_1+\delta\right)^2+\left(n_2-2\right)\left(n_1+2\delta\right) \tag{318}$$

$$E = \frac{H^2}{K^3}. \tag{319}$$

B.2 Probability Density Function

The non-central Fisher's f probability density function may be described as

$$\begin{aligned} f\left(x|n_1,n_2,\delta\right) &= -\frac{\partial}{\partial x}Q\left(x|n_1,n_2,\delta\right) \\ &\approx -\frac{\partial}{\partial x}I\left(x|\frac{n_2}{2},\frac{b}{2}\right) \\ &= -\frac{\partial y}{\partial x}\left[\frac{\partial}{\partial y}I\left(x|\frac{n_2}{2},\frac{b}{2}\right)\right]. \end{aligned} \tag{320}$$

Differentiating equations (313) and (312) yields respectively

$$\begin{aligned} \frac{\partial y}{\partial x} &= -\left(\frac{b}{n_2h}\right)\left[1+\frac{b}{n_2h}\left(x+c\right)\right]^{-2} \\ &= -\frac{by^2}{n_2h} \end{aligned} \tag{321}$$

and

$$\frac{\partial}{\partial x}I\left(x|\alpha,\beta\right) = \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}x^{\alpha-1}\left(1-x\right)^{\beta-1}. \tag{322}$$

Substitution into equation (320) results in the approximation

$$f\left(x|n_1,n_2,\delta\right) \approx \left(\frac{b}{n_2h}\right)\frac{\Gamma\left(\frac{n_2+b}{2}\right)}{\Gamma\left(\frac{n_2}{2}\right)\Gamma\left(\frac{b}{2}\right)}y^{\frac{n_2}{2}+1}\left(1-y\right)^{\frac{b}{2}-1}, \tag{323}$$

where y, b, and h are described by equations (313) - (319).

Appendix C

Statistical Distributions

For the convenience of the reader, a list of the probability density functions, along with certain properties, of the random variables encountered in this dissertation is included in this appendix. This material may be found in most mathematical statistics texts with the exception of the chi-squared and Fisher's f non-central random variables which may be found in most linear models texts. Some specific references are Manoukian [37], Muirhead [24], and Johnson and Kotz [26].

Beta $X \sim \text{Beta}(\alpha, \beta)$ $\alpha > 0$, and $\beta > 0$

PDF: $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ $0 \leq x \leq 1.$

Mean: $\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$

Variance: $\mathrm{Var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Chi-Squared $X \sim \mathcal{X}^2_n$ where n is a positive integer

PDF: $f(x) = \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$ $x > 0$

Mean: $\mathrm{E}[X] = n$

Variance: $\mathrm{Var}[X] = 2n$

MGF: $M_X(t) = (1-2t)^{-\frac{n}{2}}$ $t < \frac{1}{2}$

Complex Multivariate Normal $\mathbf{x} \sim \mathcal{CN}_n(\mu, \mathbf{\Sigma})$

PDF: $f(\mathbf{x}) = \frac{1}{\pi^n |\mathbf{\Sigma}|} \mathrm{e}^{-(\mathbf{x}-\mu)^H \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}$

Mean: $\mathrm{E}[\mathbf{x}] = \mu$

Covariance: $\mathrm{E}\left[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^H\right] = \mathbf{\Sigma}$

Complex Wishart $\mathbf{A} \sim \mathcal{CW}_n(m, \mathbf{\Sigma})$

PDF: $f(\mathbf{A}) = \frac{|\mathbf{A}|^{m-n}}{\Gamma_{m,n} |\mathbf{\Sigma}|^m} \mathrm{etr}(-\mathbf{\Sigma}^{-1}\mathbf{A})$

$\Gamma_{m,n} = \pi^{\frac{n(n-1)}{2}} \Gamma(m) \cdots \Gamma(m-n+1)$

Mean: $\mathrm{E}[\mathbf{A}] = m\mathbf{\Sigma}$

Note: If $\mathbf{x}_i \sim \mathcal{CN}_n(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{A} = \sum_{i=1}^{M} \mathbf{x}_i \mathbf{x}_i^H$ then $\mathbf{A} \sim \mathcal{CW}_n(m, \mathbf{\Sigma})$

Fisher's F $X \sim f_{n_1,n_2}$ where n_1 and n_2 are positive integers

PDF: $f(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \frac{n_2^{\frac{n_2}{2}} n_1^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}}{(n_2+n_1 x)^{\frac{n_1+n_2}{2}}}$ $x > 0$

Mean: $\mathrm{E}[X] = \frac{n_2}{n_2-2}$ $n_2 > 2$

Variance: $\mathrm{Var}[X] = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ $n_2 > 4$

Gamma $X \sim \mathrm{Gamma}(\alpha, \beta)$ where $\alpha > 0$ and $\beta > 0$

PDF: $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x}$ $x > 0$

Mean: $\mathrm{E}[X] = \frac{\alpha}{\beta}$

Variance: $\mathrm{Var}[X] = \frac{\alpha}{\beta^2}$

MGF: $M_X(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha}$ $t < \beta$

Non-Central Chi-Squared $X \sim \mathcal{X}^2_n(\delta)$ where $\delta > 0$ and n is a positive integer

PDF: $$f(x) = \sum_{k=0}^{\infty} \frac{e^{-\frac{\delta}{2}} \left(\frac{\delta}{2}\right)^k}{k!} \frac{x^{\frac{n}{2}+k-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}+k} \Gamma\left(\frac{n}{2}+k\right)} \quad x > 0$$

Mean: $$\mathrm{E}[X] = n + \delta$$

Variance: $$\mathrm{Var}[X] = 2n + 4\delta$$

MGF: $$M_X(t) = \frac{e^{-\frac{\delta}{2}} e^{\frac{\delta}{2(1-2t)}}}{(1-2t)^{\frac{n}{2}}} \qquad t < \frac{1}{2}$$

Non-Central F $X \sim f_{n_1,n_2}(\delta)$ where $\delta > 0$, and n_1 and n_2 are positive integers

PDF: $$f(x) = \sum_{k=0}^{\infty} \frac{e^{-\frac{\delta}{2}} \left(\frac{\delta}{2}\right)^k}{k!} \frac{\Gamma\left(\frac{n_1+n_2}{2}+k\right)}{\Gamma\left(\frac{n_1}{2}+k\right)\Gamma\left(\frac{n_2}{2}\right)} \frac{n_2^{\frac{n_2}{2}} n_1^{\frac{n_1}{2}+k} x^{\frac{n_1}{2}+k-1}}{(n_2+n_1 x)^{\frac{n_1+n_2}{2}+k}} \quad x > 0$$

Mean: $$\mathrm{E}[X] = \frac{n_2(n_1+\delta)}{n_1(n_2-2)} \qquad n_2 > 2$$

Variance: $$\mathrm{Var}[X] = 2\left(\frac{n_2}{n_1}\right)^2 \left[\frac{(n_1+\delta)^2 + (n_1+2\delta)(n_2-2)}{(n_2-2)^2(n_2-4)}\right] \qquad n_2 > 4$$

Bibliography

[1] E. S. Page, "Continuous Inspection Schemes," *Biometrika*, vol. 41, pp. 100–114, 1954.

[2] S. Haykin and A. O. Steinhart, eds., *Adaptive Radar Detection and Estimation.* John Wiley & Sons, 1992.

[3] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications.* Prentice-Hall, 1993.

[4] T. H. Kerr, "The Controversy over the use of SPRT and GLR Techniques and Other Loose Ends in Failure Detection," in *Proceedings of American Control Conference*, pp. 966–977, 1983.

[5] M. Basseville and A. Benveniste, eds., *Detection of Abrupt Changes in Signals and Dynamical Systems.* Springer-Verlag, 1986.

[6] B. D. Brumback and M. D. Srinath, "A Chi-Square Test for Fault Detection in Kalman Filters," *IEEE Transactions on Automatic Control*, vol. 32, no. 6, pp. 552–554, June 1987.

[7] T. H. Kerr, "Duality Between Failure Detection and Radar/Optical Maneuver Detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 4, pp. 581–584, July 1989.

[8] T. Dyson, *Topics in Nonlinear Filtering and Detection.* PhD thesis, Princeton University, 1986.

[9] B. Broder, *Quickest Detection Procedures and Transient Signal Detection.* PhD thesis, Princeton University, 1990.

[10] R. J. Stahl, "An Application of Page's Test for Detecting a Sinusoidal Disturbance of Unknown Frequency and Phase," Master's thesis, University of Connecticut, 1991.

[11] S. D. Blostein, "Quickest Detection of a Time-Varying Change in Distribution," *IEEE Transactions on Information Theory*, vol. 37, no. 4, pp. 1116–1122, July 1991.

[12] I. S. Reed, J. D. Mallet, and L. E. Brennan, "Rapid Convergence Rate in Adaptive Arrays," *IEEE Trans. on Aerospace and Electronic Systems*, vol. AES-10, no. 6, pp. 853–863, Nov. 1974.

[13] E. J. Kelly, "An Adaptive Detection Algorithm," *IEEE Trans. on Aerospace and Electronic Systems*, vol. AES-22, no. 1, pp. 115–127, March 1986.

[14] F. C. Robey, "A Covariance Modeling Approach to Adaptive Beamforming and Detection," Tech. Rep. 918, Lincoln Laboratory, July 1991.

[15] W. S. Chen and I. S. Reed, "A New CFAR Detection Test for Radar," in *Digital Signal Processing I*, pp. 198–214, Academic Press, 1991.

[16] P. K. Willett, "Signal Detection Theory." Notes from course at University of Connecticut, Spring 1992.

[17] G. Lorden, "Procedures for Reacting to a Change in Distribution," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.

[18] G. V. Moustakides, "Optimal Stopping Times for Detecting Changes in Distributions," *The Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986.

[19] Johnson and Kotz, eds., *Encylcopedia of Statistical Sciences*, vol. 1, pp. 77–81. John Wiley & Sons, 1982.

[20] L. L. Scharf, *Statistical Signal Processing.* Addison Wesley Publishing Company, 1991.

[21] K. S. Miller, *Hypothesis Testing with Complex Distributions.* Robert E. Krieger Publishing Co., 1980.

[22] Johnson and Kotz, eds., *Encylcopedia of Statistical Sciences*, vol. 4, pp. 219–225. John Wiley & Sons, 1983.

[23] N. R. Goodman, "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)," *The Annals of Mathematical Statistics*, vol. 34, pp. 152–177, March 1963.

[24] R. J. Muirhead, *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, 1982.

[25] S. R. Searle, *Linear Models.* John Wiley & Sons, 1971.

[26] Johnson and Kotz, eds., *Encylcopedia of Statistical Sciences*, vol. 6, pp. 276–284. John Wiley & Sons, 1985.

[27] M. L. Tiku, "Laguerre Series Forms of Non-central χ^2 and F Distributions," *Biometrika*, vol. 53, pp. 415–427, 1965.

[28] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities.* Cambridge University Press, 1967.

[29] N. L. Owsley and D. A. Abraham, "Preprocessing for High Resolution Beamforming," in *Proceedings of 23rd Asilomar Conf. on Signals, Systems and Computers*, 1989. Also Naval Underwater Systems Center Reprint Rpt. 8651, 3 Nov. 1989.

[30] K. A. Burgess and B. D. VanVeen, "Improved Adaptive Detection Performance via Subspace Processing," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. V–353–356, 1992.

[31] D. A. Gray, "Formulation of the maximum signal-to-noise ratio array processor in beam space," *Journal of the Acoustical Society of America*, vol. 72, no. 4, pp. 1195–1201, 1982.

[32] N. L. Owsley, "Enhanced Minimum Variance Beamforming," Tech. Rep. 8305, Naval Underwater Systems Center, Nov. 1988.

[33] T. Kailath, *Linear Systems.* Prentice-Hall, 1980.

[34] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, 1984.

[35] L. DeVroye, *Non-Uniform Random Variate Generation.* Springer-Verlag, 1986.

[36] R. L. Streit, "An Upper Bound on Feature Vector Dimension as a Function of Design Set Size for Two Gaussian Populations," Tech. Memo. 921048, Naval Undersea Warfare Center, March 1992.

[37] E. B. Manoukian, *Modern Concepts and Theorems of Mathematical Statistics.* Springer-Verlag, 1985.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Coast Guard Academy (J. Wolcin)	1
Office of Naval Research (Code 451: T. G. Goldsberry, N. Harned, M. Shipley)	3
Program Executive Officer, USW ASTO (J. Polcari)	1
Space and Naval Warfare Systems Command (R. Holland)	1
Naval Undersea Warfare Center, Detachment West Palm Beach (R. Kennedy)	1
Pennsylvania State University (R. Young)	1
Princeton University (S. Schwartz)	1
Defense Technical Information Center	2